

The One-Bit Null Space Learning Algorithm and its Convergence

Yair Noam and Andrea J. Goldsmith, *Fellow, IEEE*

Abstract

This paper proposes a new algorithm for MIMO cognitive radio Secondary Users (SU) to learn the null space of the interference channel to the Primary User (PU) without burdening the PU with any knowledge or explicit cooperation with the SU. The knowledge of this null space enables the SU to transmit in the same band simultaneously with the PU by utilizing separate spatial dimensions than the PU. Specifically, the SU transmits in the null space of the interference channel to the PU. We present a new algorithm, called the One-Bit Null Space Learning Algorithm (OBNSLA), in which the SU learns the PU's null space by observing a binary function that indicates whether the interference it inflicts on the PU has increased or decreased in comparison to the SU's previous transmitted signal. This function is obtained by listening to the PU transmitted signal or control channel and extracting information from it about whether the PU's Signal to Interference plus Noise power Ratio (SINR) has increased or decreased. The SU iteratively modifies its null space estimate and the corresponding interference it inflicts on the PU and measures the effect this modification has on the PU's SINR in order to refine its null space estimate.

In addition to introducing the OBNSLA, this paper provides a thorough convergence analysis of this algorithm. The OBNSLA is shown to have a linear convergence rate and an asymptotic quadratic convergence rate. Finally, we derive bounds on the interference that the SU inflicts on the PU as a function of a parameter determined by the SU. This lets the SU control the maximum level of interference, which enables it to protect the PU completely blindly with minimum complexity. The asymptotic analysis and the derived bounds also apply to the recently proposed Blind Null Space Learning Algorithm.

This work is supported by the ONR under grant N000140910072P00006, the AFOSR under grant FA9550-08-1-0480, and the DTRA under grant HDTRA1-08-1-0010. The authors are with the Dept. of Electrical Engineering, Stanford University, Stanford CA, 940305.

I. INTRODUCTION

The emergence of Multiple Input Multiple Output (MIMO) communication opens new directions in Cognitive Radio (CR) networks [1–3]. In particular, in underlay CR networks, MIMO technology enables the Secondary User (SU) to transmit a significant amount of power simultaneously in the same band as the Primary User (PU) without interfering with it, if the SU utilizes separate spatial dimensions than the PU. This spatial separation requires that the interference channel from the SU to the PU be known to the SU. Thus, acquiring this knowledge, or operating without it, is a major topic of active research, as discussed in [4]. We consider MIMO primary and secondary systems defined as follows: we assume a flat-fading MIMO channel with one PU and one SU, as depicted in Fig. 1. Let \mathbf{H}_{ps} be the channel matrix between the SU's transmitter and the PU's receiver, hereafter referred to as the SU-Tx and PU-Rx, respectively. In the underlay CR paradigm, SUs are constrained not to inflict “harmful” interference on the PU-Rx. This can be achieved if the SU restricts its signal to lie within the null space of \mathbf{H}_{ps} ; however, this is only possible if the SU knows \mathbf{H}_{ps} . Yi [5], Chen et. al. [6], and Gao et. al. [7] proposed blind solutions to this problem where the SU learns the channel matrix based on channel reciprocity: specifically, where the SU listens to the PU's transmitted signal and estimates \mathbf{H}_{ps} 's null space from the signal's second order statistics. Since these works require channel reciprocity, they are restricted to PUs that use Time Division Duplexing (TDD).

Unless there is channel reciprocity, obtaining \mathbf{H}_{ps} by the SU requires cooperation with the PU in the estimation phase; e.g. where the SU transmits a training sequence, from which the PU estimates \mathbf{H}_{ps} and feeds it back to the SU. Cooperation of this nature increases the system complexity overhead, since it requires a handshake between both systems and, in addition, the PU needs to be synchronized with the SU's training sequence. In [4], we proposed the Blind Null Space Learning Algorithm (BNSLA), which enables a MIMO underlay CR to learn the null space of \mathbf{H}_{ps} without learning the full matrix. Furthermore, during this learning, the PU does not cooperate at all with the SU and operates as though there were no other systems in the medium (the way current PUs operate today). However, the BNSLA requires the following observation constraint: at each short training interval, the SU must observe some monotone continuous function of the PU's Signal to Noise plus Interference power Ratio (SINR). For example, if the PU is using continuous power control, the PU's signal power is a monotone function of its SINR.

This paper makes two contributions. The first contribution is a new algorithm, called the One-Bit Null Space Learning Algorithm (OBNSLA), which requires much less information than the BNSLA; namely,

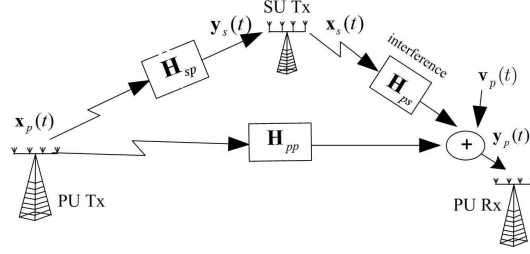


Fig. 1. Our cognitive radio scheme. H_{ps} is unknown to the secondary transmitter and $v_p(t)$ is a stationary noise (which may include stationary interference). The interference from the SU, $H_{ps}x_s(t)$, is treated as noise; i.e., there is no interference cancellation.

the SU can infer whether the interference it inflicts on the PU has increased or decreased compared to a previous time interval with a one-bit function. In other words, in the OBNSLA the SU measures a one-bit function of the PU's SINR, rather than a continuous-valued function as in the BNSLA. Using this single bit of information, the SU learns H_{ps} 's null space by iteratively modifying the spatial orientation of its transmitted signal and measuring the effect of this modification on the PU's SINR. The second contribution of the paper is to provide a thorough convergence analysis of the OBNSLA. We show that the algorithm converges linearly and has an asymptotically quadratic convergence rate. In addition, we derive upper bounds on the interference that the SU inflicts on the PU; these results enable the SU control the interference to the PU without any cooperation on its part. Furthermore, all the bounds and the convergence results apply equally to the BNSLA.

The remainder of this paper is organized as follows: Section II provides the system description and notation. Section III presents the One-Bit Null Space Learning Algorithm. Section IV provides a convergence analysis of the algorithm and presents bounds on the interference that the SU inflicts on the PU. Simulations and conclusions are presented in Sections VI and VII, respectively.

II. THE ONE-BIT NULL SPACE LEARNING PROBLEM

Consider a flat fading MIMO interference channel with a single PU and a single SU without interference cancellation; i.e., each system treats the other system's signal as noise. The PU's received signal is given by

$$y_p(t) = H_{pp}x_p(t) + H_{ps}x_s(t) + v_p(t), \quad t \in \mathbb{N} \quad (1)$$

where x_p , x_s is the PU's and SU's transmitted signal, respectively, H_{pp} is the PU's direct channel; H_{ps} is the interference channel between the PU Rx and the SU Tx, and $v_p(t)$ is a zero mean stationary noise. In the underlay CR paradigm, the SU is constrained not to exceed a maximum interference level at the

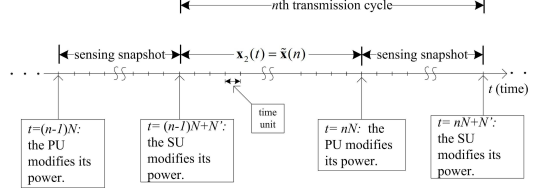


Fig. 2. The time indexing used in this paper. t indexes the basic time unit (pulse time) where N time units constitute a TC that is indexed by n . Furthermore, K transmission cycles constitute a learning phase (not shown in this figure).

PU Rx; i.e.,

$$\|\mathbf{H}_{ps}\mathbf{x}_s(t)\|^2 \leq \eta_{\max}, \quad (2)$$

where $\eta_{\max} > 0$ is the maximum interference constraint. If $\eta_{\max} = 0$, the SU is strictly constrained to transmit only within the null space of the matrix \mathbf{H}_{ps} . In this paper, all vectors are column vectors. Let \mathbf{A} be an $l \times m$ complex matrix; then, its null space is defined as $\mathcal{N}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{C}^m : \mathbf{A}\mathbf{y} = \mathbf{0}\}$ where $\mathbf{0} = [0, \dots, 0]^T \in \mathbb{C}^l$.

Since our focus is on constraining the interference caused by the SU to the PU, we only consider the term $\mathbf{H}_{ps}\mathbf{x}_s(t)$ in (1). Hence, \mathbf{H}_{ps} and \mathbf{x}_s will be denoted by \mathbf{H} and \mathbf{x} , respectively. We also define the Hermitian matrix \mathbf{G} as

$$\mathbf{G} = \mathbf{H}^* \mathbf{H} \quad (3)$$

The time line \mathbb{N} is divided into N -length intervals, each referred to as a transmission cycle (TC), as depicted in Figure 2. For each TC, the SU's signal is constant; i.e.,

$$\mathbf{x}_s((n-1)N + N') = \mathbf{x}_s((n-1)N + 1) = \dots = \mathbf{x}_s(nN + N' - 1) \triangleq \tilde{\mathbf{x}}(n), \quad (4)$$

where the time interval $nN < t \leq nN + N' - 1$ is the snapshot in which the SU measures a function $q(n)$ that satisfies the following observation constraint.

Observation Constraint (OC) on the function $q(n)$: There exists some integer $M \geq 1$, such that given $q(n-m), \dots, q(n)$, the SU can extract

$$\tilde{h}(\tilde{\mathbf{x}}(n), \tilde{\mathbf{x}}(n-m)) = \begin{cases} 1, & \text{if } \|\mathbf{H}\tilde{\mathbf{x}}(n)\| \geq \|\mathbf{H}\tilde{\mathbf{x}}(n-m)\| \\ -1, & \text{otherwise} \end{cases}, \quad (5)$$

for every $m \leq M$.

The SU's objective is to learn $\mathcal{N}(\mathbf{H})$ from $\{\tilde{\mathbf{x}}(n), q(n)\}_{n \in \mathbb{N}}$. This problem, referred to as the One-bit Blind Null Space Learning (OBNSL) problem, is illustrated in Figure 3 for $M = 1$. The OBNSL problem

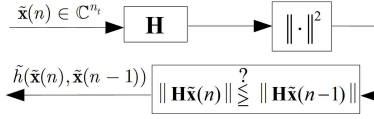


Fig. 3. Block Diagram of the One-bit Blind Null Space Learning Problem. The SU's objective is to learn the null space of \mathbf{H} by inserting a series of $\{\tilde{\mathbf{x}}(n)\}_{n \in \mathbb{N}}$ and measuring $\tilde{h}(\tilde{\mathbf{x}}(n), \tilde{\mathbf{x}}(n-1))$ as output.

is similar to the Blind Null Space Learning (BNSL) problem [4] except for one important difference. In the latter, the SU observes a continuous-valued function of the PU's SINR whereas in the OBNSL problem, it observes a one-bit function. In both problems, the SU obtains $q(n)$ by measuring the PU's transmit energy, or any other parameter that indicates the PU's SINR (see Sec. II-B in [4] for examples). However, in the OBNSL problem, the SU is more flexible since it can obtain $q(n)$ from, for example, incremental power control¹ or other quantized functions of the PU's SINR such as modulation size. From a system point of view, the OC means that between m consecutive transmission cycles, the PU's SINR is only affected by variations in the SU's signal. The learning process unfolds as follows. In the first TC ($n = 1$), the SU transmits $\tilde{\mathbf{x}}(1)$, and measures $q(n)$. In the next TC, the SU transmits $\tilde{\mathbf{x}}(2)$ and measures $q(2)$ from which it extracts $\tilde{h}(\tilde{\mathbf{x}}(1), \tilde{\mathbf{x}}(2))$. This process is repeated until the null space is approximated. Note that while $\tilde{h}(\tilde{\mathbf{x}}(1), \tilde{\mathbf{x}}(2))$ requires two TCs, $\tilde{h}(\tilde{\mathbf{x}}(n-1), \tilde{\mathbf{x}}(n))$ for $n > 2$ requires a single TC. In the following section we describe the algorithm that performs this learning. We will also discuss the role of M in the OC. This parameter, which indicates the "memory" that $q(n)$ has of the PU's SINR, may affect the number of TCs required for learning $\mathcal{N}(\mathbf{H})$; i.e., a larger M reduces the number of TCs.

III. THE ONE-BIT BLIND NULL SPACE LEARNING ALGORITHM (OBNSLA)

We now present the OBNSLA by which the SU approximates $\mathcal{N}(\mathbf{H})$ from $\{\tilde{\mathbf{x}}(n), q(n)\}_{n=1}^T$ under the OC, where the approximation error can be made arbitrarily small for sufficiently large T . Once the SU learns $\mathcal{N}(\mathbf{H})$, it can optimize its transmitted signal, regardless of the optimization criterion, under the constraint that its signal lies in $\mathcal{N}(\mathbf{H})$. Let $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ be \mathbf{H} 's Singular Value Decomposition (SVD), where \mathbf{V} and \mathbf{U} are $n_t \times n_t$ and $n_r \times n_r$ unitary matrices, respectively, and assume that $n_t > n_r$. The matrix $\mathbf{\Sigma}$ is an $n_r \times n_t$ diagonal matrix with real nonnegative diagonal entries $\sigma_1, \dots, \sigma_d$ arranged as

¹This is power control that is carried out using one-bit command which indicates whether to increase or decrease the power by a certain amount.

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_d > 0$. We assume without loss of generality that $n_r = d (= \text{Rank}(\mathbf{H}))$. In this case

$$\mathcal{N}(\mathbf{H}) = \text{span}(\mathbf{v}_{n_r+1}, \dots, \mathbf{v}_{n_t}), \quad (6)$$

where \mathbf{v}_i denotes \mathbf{V} 's i th column. From the SU's point of view, it is sufficient to learn $\mathcal{N}(\mathbf{G})$ (recall, $\mathbf{G} = \mathbf{H}^* \mathbf{H}$), which is equal to $\mathcal{N}(\mathbf{H})$ since

$$\mathbf{G} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^*, \quad (7)$$

where $\mathbf{\Lambda} = \mathbf{\Sigma}^T \mathbf{\Sigma}$. The decomposition in (7) is known as the Eigenvalue Decomposition (EVD) of \mathbf{G} . In order to obtain $\mathcal{N}(\mathbf{H})$ it is sufficient to obtain \mathbf{G} 's EVD. However, in the OBNSL problem, \mathbf{G} is not observed, so the SU needs to obtain the EVD using only one-bit information. The OBNSL algorithm does so by blindly implementing the well-known Cyclic Jacobi Technique (CJT) for Hermitian matrix diagonalization. We begin with a review of that technique.

A. Review of the Cyclic Jacobi Technique

The Jacobi technique [see e.g. 8] obtains the EVD of the $n_t \times n_t$ Hermitian matrix \mathbf{G} via a series of 2-dimensional rotations that eliminates two off-diagonal elements at each step (indexed by k). It begins by setting $\mathbf{A}_0 = \mathbf{G}$ and then performs the following rotation operations $\mathbf{A}_{k+1} = \mathbf{V}_k \mathbf{A}_k \mathbf{V}_k^*$, $k = 1, 2, \dots$, where

$$\mathbf{V}_k = \mathbf{R}_{l,m}(\theta, \phi) \quad (8)$$

is an $n_t \times n_t$ unitary rotation matrix that is equal to \mathbf{I}_{n_t} except for its m th and l th diagonal entries that are equal to $\cos(\theta)$, and its (m, l) th and (l, m) th entries that are equal to $e^{-i\phi} \sin(\theta)$ and $-e^{i\phi} \sin(\theta)$, respectively. For each k , the values of θ, ϕ are chosen such that $[\mathbf{A}_k]_{l,m} = 0$, or stated differently, θ and ϕ are chosen to zero the l, m and m, l off diagonal entries of \mathbf{A}_k (which are conjugate to each other). Note that in an $n_t \times n_t$ Hermitian matrix, there are $(n_t - 1)n_t/2$ such pairs. The values of l, m are chosen in step k according to a function $J : \mathbb{N} \rightarrow \{1, \dots, n_t\} \times \{1, \dots, n_t\}$, i.e $J_k = (l_k, m_k)$. It is the choice of J_k that differs between different Jacobi techniques. In the cyclic Jacobi technique, l_k, m_k satisfy $1 < l_k < n_t - 1$ and $l_k < m_k \leq n_t$ such that each pair (l, m) is chosen once in every $(n_t - 1)n_t/2$ rotations. Such $(n_t - 1)n_t/2$ rotations are referred to as a Jacobi sweep. An example of a single sweep of the CJT for $n_t = 3$ is the following series of rotations: $J_1 = (1, 2)$, $J_2 = (1, 3)$, $J_3 = (2, 3)$. The next sweep is $J_4 = (1, 2)$, $J_5 = (1, 3)$, $J_6 = (2, 3)$ and so forth.

The convergence of the CJT has been studied extensively over the last sixty years. The first proof of

convergence of the CJT for complex Hermitian matrices was given by Foster and Henrici [9]. However, this result did not determine the convergence rate. The convergence rate problem was addressed by Henrici and Zimmermann [10], who proved that the CJT for real symmetric matrices has a global linear convergence rate² that depends on the matrix size n_t if $\theta_k \in [-\pi/4, \pi/4]$ for every k . Fernando [11] extended this result to complex Hermitian matrices. It was later shown by Henrici [12], and by Wilkinson [13] that in the case of a complex Hermitian matrix with well separated eigenvalues, the CJT has a quadratic convergence³ rate. Hari [14] extended the results of Henrici and Wilkinson by proving a quadratic convergence rate under more general conditions, including identical eigenvalues and clusters of eigenvalues (that is, very close eigenvalues). Studies have shown that in practice the number of iterations that is required for the CJT to reach its asymptotic quadratic convergence rate is a small number, but this has not been proven rigorously. Brent and Luk [15] have argued heuristically that this number is $O(\log_2(n_t))$ cycles for $n_t \times n_t$ matrices. Extensive numerical results show that quadratic convergence is obtained after three to four cycles (see e.g. [8, page 429], [16, page 197]). Thus, since each Jacobi sweep has $n_t(n_t - 1)/2$ rotations, the overall number of rotations in the CJT roughly grows as n_t^2 . For further details about the CJT and its convergence, the reader is referred to [8, 16, 17].

B. The One-Bit Line Search

The learning in the OBNSLA is carried out in learning stages, indexed by k , where each stage performs one Jacobi rotation. The SU approximates the matrix \mathbf{V} by \mathbf{W}_{k_s} , where

$$\mathbf{W}_k = \mathbf{W}_{k-1} \mathbf{R}_{l,m}(\theta_k, \phi_k), \quad k = 1, \dots, k_s, \quad (9)$$

and $\mathbf{W}_0 = \mathbf{I}$. Recall that in the Cyclic Jacobi technique, one observes the matrix

$$\mathbf{A}_{k-1} = \mathbf{W}_{k-1}^* \mathbf{G} \mathbf{W}_{k-1} \quad (10)$$

and chooses the rotation angles which zero \mathbf{A}_k 's (l, m) th off diagonal entry; i.e.,

$$[\mathbf{R}_{l,m}^*(\theta_k^J, \phi_k^J) \mathbf{A}_{k-1} \mathbf{R}_{l,m}(\theta_k^J, \phi_k^J)]_{l,m} = 0 \quad (11)$$

²A sequence a_n is said to have a linear convergence rate of $0 < \beta < 1$ if there exists $n_0 \in \mathbb{N}$ such that $|a_{n+1}| < \beta|a_n|$ for every $n > n_0$. If $n_0 = 1$, a_n has a global linear convergence rate.

³A sequence is said to have a quadratic convergence rate if there exists $\beta > 0, n_0 \in \mathbb{N}$ such that $|a_{n+1}| < \beta|a_n|^2, \forall n > n_0$.

In the OBNSL problem, the SU needs to perform this step using only $\{\tilde{\mathbf{x}}(n), q(n)\}_{n=1}$, without observing the matrix \mathbf{A}_{k-1} . The following theorem, proved in [4], is the first step towards such a blind implementation of the Jacobi technique. The theorem converts the problem of obtaining the optimal Jacobi rotation angles into two one-dimensional optimizations of the function $S(\mathbf{A}_{k-1}, \mathbf{r}_{l,m}(\theta, \phi))$ (which is continuous, as shown in [4]), where $S(\mathbf{A}, \mathbf{x}) = \mathbf{x}^* \mathbf{A} \mathbf{x}$ and $\mathbf{r}_{l,m}(\theta, \phi)$ is $\mathbf{R}_{l,m}(\theta, \phi)$'s l th column.

Theorem 1: [4, Theorem 2] Consider a $n_t \times n_t$ Hermitian matrix \mathbf{A}_{k-1} in (10), and let $S(\mathbf{A}, \mathbf{x}) = \mathbf{x}^* \mathbf{A} \mathbf{x}$ and $\mathbf{r}_{l,m}(\theta, \phi)$ be $\mathbf{R}_{l,m}(\theta, \phi)$'s l th column. The optimal Jacobi parameters θ_k^J and ϕ_k^J , which zero out the (l, m) th entry of $\mathbf{R}_{l,m}^*(\theta_k^J, \phi_k^J) \mathbf{A}_{k-1} \mathbf{R}_{l,m}(\theta_k^J, \phi_k^J)$, are given by

$$\phi_k^J = \arg \min_{\phi \in [-\pi, \pi]} S(\mathbf{A}_{k-1}, \mathbf{r}_{l,m}(\pi/4, \phi)) \quad (12)$$

$$\theta_k^J = T_k(\phi_k^J) \quad (13)$$

where

$$T_k(\phi) = \begin{cases} \tilde{\theta}_k(\phi) & \text{if } -\frac{\pi}{4} \leq \tilde{\theta}_k(\phi) \leq \frac{\pi}{4} \\ \tilde{\theta}_k(\phi) - \text{sign}(\tilde{\theta}_k(\phi))\pi/2 & \text{otherwise} \end{cases} \quad (14)$$

where $\text{sign}(x) = 1$ if $x > 0$ and -1 otherwise, and

$$\tilde{\theta}_k(\phi) = \arg \min_{\theta \in [-\pi/2, \pi/2]} S(\mathbf{A}_{k-1}, \mathbf{r}_{l,m}(\theta, \phi)) \quad (15)$$

The theorem enables the SU to solve the optimization problems in (12) and (15) via line searches based on $\{\tilde{\mathbf{x}}(n), q(n)\}_{n=1}^T$. This is because under the OC, the SU can extract $\tilde{h}(\tilde{\mathbf{x}}(n), \tilde{\mathbf{x}}(n-m))$, which indicates whether $S(\mathbf{G}, \tilde{\mathbf{x}}(n)) > S(\mathbf{G}, \tilde{\mathbf{x}}(n-m))$ is true or false. It is possible, however, to further reduce the complexity of the line search, which is important, since each search point requires a TC. To see this, consider the line search in (12) and denote $w(\phi) = S(\mathbf{A}_k, \mathbf{r}_{l,m}(\pi/4, \phi)) = \|\mathbf{H}\mathbf{W}_{k-1}\mathbf{r}_{l,m}(\pi/4, \phi)\|^2$. According to the OC, for each ϕ_1, ϕ_2 the SU only knows whether $w(\phi_1) \geq w(\phi_2)$ or not. Assume that the SU tries to approximate ϕ_k^J by searching over a linear grid, with a spacing of η , on the interval $[-\pi, \pi]$. The complexity of such a search is at least $O(1/\eta)$ since each point in the grid must be compared to a different point at least once. The two line searches in (12) and (15) would be carried out much more efficiently if binary searches could be invoked. However, a binary search is feasible only if the objective function has a unique local minimum point, which is not the case in (12) and (15) because

$$S(\mathbf{G}, \mathbf{r}_{l,m}(\theta, \phi)) = \cos^2(\theta) |g_{l,l}| + \sin^2(\theta) |g_{m,m}| - |g_{l,m}| \sin(2\theta) \cos(\phi + \angle g_{l,m}) \quad (16)$$

Thus, before invoking the binary search, a single-minimum interval (SMI) must be determined; i.e., an interval in which the target function in (12) or (15) has a single local minimum. This is possible via the following proposition:

Proposition 2: Let $w(\phi) = \|\mathbf{H}\mathbf{r}_{l,m}(\pi/4, \phi)\|^2$ where $\mathbf{r}_{l,m}$ is defined in Theorem 1. Let $\check{\phi} \in [-\pi, \pi]$ be a minimum⁴ point of $w(\phi)$, then

- (a) $\check{\phi} \in [-3\pi/4, -\pi/4]$ if $w(-\pi), w(0) \geq w(-\pi/2)$.
- (b) $\check{\phi} \in [-\pi/4, \pi/4]$ if $w(-\pi) \geq w(-\pi/2) \geq w(0)$.
- (c) $\check{\phi} \in [\pi/4, 3\pi/4]$ if $w(-\pi), w(0) \leq w(-\pi/2)$.
- (d) $\check{\phi} \in [3\pi/4, \pi] \cup [-\pi, -3\pi/4]$ if $w(-\pi) \leq w(-\pi/2) \leq w(0)$.

Proof: From (16), $w(\phi)$ can be expressed as $w(\phi) = B - A \cos(\phi - \check{\phi})$, $A, B \geq 0$. If $B = 0$, every $\phi \in [-\pi, \pi]$ is a minimum point and the proposition is true. We now assume that $B > 0$. By substituting $w(0) = B - A \cos(\check{\phi})$, $w(\pi/2) = B - A \sin(\check{\phi})$ and $w(\pi) = B + A \cos(\check{\phi})$ into $w(-\pi), w(0) \geq w(-\pi/2)$, one obtains that the latter is equivalent to $\pm \cos(\check{\phi}) > \sin(\check{\phi})$ which is equivalent to $(-3\pi/4 < \check{\phi} < \pi/4) \cap ((-\pi < \check{\phi} < -\pi/4) \cup (3\pi/4 < \check{\phi} < \pi))$. The last set can be written as $\check{\phi} \in (-3\pi/4, -\pi/4)$, which establishes (a). The proof of (b)-(d) is similar. \square

Note that unless w is a horizontal line, it is a 2π periodical sinusoid. In the latter case, there cannot be more than a single local (and therefor global) minimum within an interval of $\pi/2$. If w is a horizontal line, every point is a minimum point. In both cases, a binary search can be invoked; i.e., the SU can efficiently approximate θ_k^J, ϕ_k^J by $\hat{\theta}_k^J$ and $\hat{\phi}_k^J$, respectively, using a binary search, such that

$$|\hat{\theta}_k^J - T_k(\hat{\phi}_k^J)|, |\hat{\phi}_k^J - \phi_k^J| \leq \eta, \quad (17)$$

where $\eta > 0$ determines the approximation accuracy. In order to invoke a binary search, the SU uses Proposition 2, to determine an SMI via $u_n(\pi, \pi/2)$ and $u_n(\pi/2, 0)$, where $u_n(\phi_n, \phi_{n-1}) = \tilde{h}_n(\mathbf{r}_{l,m}(\pi/4, \phi_n), \mathbf{r}_{l,m}(\pi/4, \phi_{n-1}))$, and \tilde{h} is defined in (5). The one-bit line search is given in Algorithm 1. In determining the SMI, the one-bit line search requires 3 TCs: two TCs for $u_n(\pi, \pi/2)$, and one more for $u_n(\pi, 0)$. Given an SMI of length a and an accuracy of $\eta > 0$, it takes $\lfloor -\log_2(\eta/a) \rfloor + 1$ search points to obtain the minimum to within that accuracy. In the search for ϕ_k^J , $a = \pi/2$, thus $\hat{\phi}_k^J \in [\phi_k^J - \eta, \phi_k^J + \eta]$ is obtained using $\lfloor -\log(2\eta/\pi) \rfloor + 1$ TCs, plus the 3 TCs required for determining the SMI and 2 more TCs to compare the initial boundaries of the SMI. Then, θ_k^J can be approximated to within η in the same way.

⁴Since w is a 2π periodical sinusoid, such a point always exists, though might not be unique.

We conclude with a discussion of parameter M in the OC. The proposed line search can obtain $w(\phi)$'s minimum even for $M = 1$. However, the number of TCs is lower if M is larger. Assume that the SU has obtained an SMI $[\phi^{\min}, \phi^{\max}]$. It takes the SU two TCs where, in the first, it transmits $r_{l,m}(\pi/4, \phi_1)$, where $\phi_1 = \phi^{\min}$ and measures $q(1)$, and in the second TC, it transmits $r_{l,m}(\pi/4, \phi_2)$, where $\phi_2 = \phi^{\max}$ and measures $q(2)$. These q 's are sufficient for determining whether

$$w(\phi_{\max}) > w(\phi_{\min}) \quad (18)$$

If (18) is false, ϕ_{\max} is set as $\phi_{\max} = (\phi_{\min} + \phi_{\max})/2$. In the next phase of the binary search the SU transmits $r_{l,m}(\pi/4, \phi_3)$, where $\phi_3 = \phi_{\max}$ and measures $q(3)$. However, because $M = 1$ it cannot use $q(1)$ to check (18); it needs an additional TC to do so. This extra TC is not required if $M > 1$. In general, $M > \lfloor -\log(\eta/\pi) + 1 \rfloor$ (which is the maximum number of search points required for the binary search) guarantees that each search point requires one TC.

C. The OBNSLA

Now that we have established the one-bit line search, we can present the OBNSLA. In the OBNSLA, the SU performs two line searches for each k . The first search is carried out to find $\hat{\phi}_k$ that minimizes $\|\mathbf{H}\mathbf{W}_k \mathbf{r}_{l_k, m_k}(\pi/4, \phi)\|^2$, where each search point, ϕ_n , is obtained by one TC in which the SU transmits

$$\mathbf{x}_s(t) = \tilde{\mathbf{x}}(n) = \mathbf{W}_k \mathbf{r}_{l_k, m_k}(\pi/4, \phi_n) \in \mathbb{C}^{n_t}, \quad \forall (n-1)N \leq t \leq nN, \quad (19)$$

and measures $q(n)$. In first line search, the SU obtains $\hat{\phi}_k^J$ which is then used in the second line search to obtain $\tilde{\theta}_k(\hat{\phi}_k^J)$ according to (15), and then to obtain $\hat{\theta}_k^J$ according to (13). The indices (l_k, m_k) are chosen as in the CJT.

After performing k_s iterations the SU approximates the matrix \mathbf{V} (see (7)) by \mathbf{W}_{k_s} . It then chooses its pre-coding matrix \mathbf{T}_{k_s} as

$$\mathbf{T}_{k_s} = [\mathbf{w}_{i_1}^{k_s}, \dots, \mathbf{w}_{i_{n_t-n_r}}^{k_s}], \quad (20)$$

where \mathbf{w}_i^k is \mathbf{W}_k 's i th column, and $i_1, i_2, \dots, i_{n_t-n_r}$ is an indexing such that $(\mathbf{w}_{i_q}^{k_s})^* \mathbf{G} \mathbf{w}_{i_q}^s \leq (\mathbf{w}_{i_v}^{k_s})^* \mathbf{G} \mathbf{w}_{i_v}^s$ for every $q \leq v$. Thus, the interference power that the SU inflicts on the PU is bounded as $\|\mathbf{H}\mathbf{x}\|^2 \leq p_s \|\mathbf{H}\mathbf{w}_{i_{n_t-n_r}}^{k_s}\|^2, \forall \|\mathbf{x}\|^2 = p_s$, where p_s is the SU's transmit power. The OBNSLA algorithm is given in Algorithm 2. It is important to note that since only $\tilde{h}_n(\tilde{\mathbf{x}}(n), \tilde{\mathbf{x}}(n-1))$ is observed, rather than

$\|\mathbf{H}\mathbf{x}(n-1)\|^2$ and $\|\mathbf{H}\mathbf{x}(n)\|^2$, the eigenvalues of \mathbf{G} cannot be obtained by the OBNSLA⁵.

Although the SU becomes “invisible” to the PU after it learns $\mathcal{N}(\mathbf{H}_{ps})$, it interferes with the PU during this learning process. Furthermore, this interference is an important ingredient in the learning since it provides the SU with the means to learn $\mathcal{N}(\mathbf{H}_{ps})$, i.e. $q(n)$. Nevertheless, the SU must also protect the PU during the learning process. Hence we assume that there exists an additional mechanism enabling the SU to choose $\tilde{\mathbf{x}}(n)$ ’s power to be high enough to be able to extract $q(n)$, but not too high, so as to meet the interference constraint (2). We give examples for such mechanisms in [4, Sec. II-C].

Algorithm 1 $[z, n] = \text{OneBitLineSearch}(\{\tilde{h}_l\}_{l \in \mathbb{N}}, z_{\max}, n, \eta, \mathbf{x}(z))$

Initialize: $L \leftarrow z_{\max}$,
 $u_n(z_1, z_2) \leftarrow \tilde{h}_n(\mathbf{x}(z_1), \mathbf{x}(z_2))$
 $a \leftarrow u_n(-L, -L/2), n++$.
 $b \leftarrow u_n(0, -L/2), n++$
 $z_{\max} \leftarrow (3 + 2b - 2a(1 + 2b))L/4$;
 $z_{\min} \leftarrow z_{\max} - L/2$.
while $|z_{\max} - z_{\min}| \geq \eta$ **do**
 $z \leftarrow (z_{\max} + z_{\min})/2$
 $a \leftarrow u(z_{\max}, z_{\min}), n++$
 if $a=1$ **then**
 $z_{\max} \leftarrow z$
 else
 $z_{\min} \leftarrow z$
 end if
end while

IV. ALGORITHM CONVERGENCE

The OBNSLA is, in fact, a blind implementation of the CJT whose convergence properties have been extensively studied over the last 60 years. However, the convergence results of the CJT do not apply directly to the OBNSLA. This is because of the approximation in (17); i.e., due to the fact that for every k , the rotation angles $\hat{\theta}_k^J, \hat{\phi}_k^J$ are obtained by a binary search of accuracy η . Thus the off diagonal entries are not completely annihilated; i.e., $[\mathbf{A}_{k+1}]_{l_k, m_k} \approx 0$ instead of $[\mathbf{A}_{k+1}]_{l_k, m_k} = 0$. Moreover, we would like to make this line search accuracy as low as possible (that is, to make η as large as possible) in order to reduce the number of TCs. It is therefore crucial to understand how η affects the performance of the OBNSLA algorithm, in terms of convergence rate and the interference reduction to the PU. In this

⁵In [18] it is shown that if $S(\mathbf{G}, \mathbf{x})$ is known, the problem can be simplified drastically where \mathbf{G} can be obtained precisely by a finite number of TCs.

Algorithm 2 The OBNSL Algorithm

Input: $\{\tilde{h}_v\}_{v \in \mathbb{N}}$, defined in (5).

Output: \mathbf{W}

initialize: $n = 1$

$[\mathbf{W}, n] = \text{OBNSLF}(\{\tilde{h}_v\}_{v \in \mathbb{N}}, n_t, n)$

End

Function: $[\mathbf{W}, n] = \text{OBNSLF}(\{\tilde{h}_v\}_{v \in \mathbb{N}}, n_t, n)$

Initialize: $k = 1, \mathbf{W} = \mathbf{I}_{n_t}, \Delta_j = 2\eta, \forall j \leq 0$

while $(\max_{j \in \{k-n_t(n_t-1)/2, \dots, k\}} \Delta_j \geq \eta)$ **do**

$\mathbf{x}(\phi) \leftarrow \mathbf{W}\mathbf{r}_{l_k, m_k}(\pi/4, \phi)$

$[\hat{\phi}_k, n] \leftarrow \text{OneBitLineSearch}(\{\tilde{h}_l\}_{l \in \mathbb{N}}, \pi, n, \eta, \mathbf{x}(\phi))$

$\mathbf{x}(\theta) \leftarrow \mathbf{W}\mathbf{r}_{l_k, m_k}(\theta, \hat{\phi}_k)$

$[\tilde{\theta}_k, n] \leftarrow \text{OneBitLineSearch}(\{\tilde{h}_l\}_{l \in \mathbb{N}}, \pi/2, n, \eta, \mathbf{x}(\theta))$

$\hat{\theta}_k \leftarrow \tilde{\theta}_k$ if $\tilde{\theta}_k \leq |\pi/4|$, otherwise $\hat{\theta}_k \leftarrow \tilde{\theta}_k - \pi \text{sign}(\tilde{\theta}_k)/2$.

$\Delta_k \leftarrow |\theta_k|$

$\mathbf{W} \leftarrow \mathbf{W}\mathbf{R}_{l_k, m_k}(\hat{\theta}_k, \hat{\phi}_k)$

$k \leftarrow k + 1$.

end while

section, we extend the classic convergence results of the CJT to the OBNSLA and indicate the required accuracy in the binary search that assures convergence and bounds the maximum reduction level of the interference inflicted by the SU on the PU. It will also be shown that the same convergence analysis applies to the BNSLA proposed in [4].

A. Global Linear Convergence Rate

The following theorem shows that for a sufficiently good line search accuracy, the OBNSLA has a global linear convergence rate.

Theorem 3: Let \mathbf{G} be a finite dimensional $n_t \times n_t$ complex Hermitian matrix and P_k denotes the Frobenius norm of the off diagonal upper triangular (or lower triangular) part of $\mathbf{A}_k = \mathbf{W}_{k-1}^* \mathbf{G} \mathbf{W}_{k-1}$ where \mathbf{W}_k is defined in (9) and let $m = n_t(n_t - 1)/2$. Let η be the accuracy of the binary search (see (17)), then the OBNSLA satisfies

$$P_{k+m}^2 \leq P_k^2 (1 - 2^{-(n_t-2)(n_t-1)/2}) + (n_t^2 - n_t)(7 + 2\sqrt{2})\eta^2 \|\mathbf{G}\|^2 \quad (21)$$

Proof: See Appendix A.

Comment: For $\eta = 0$, we obtain the well known convergence result by Fernando [11].

B. An Asymptotic Quadratic Convergence Rate

So far, it has been shown that for the right choice of η , the OBNSLA converges and that for sufficiently small η it has an approximately global linear convergence rate. In what follows, it will be shown that for sufficiently small η , the OBNSLA has an asymptotic quadratic convergence rate, but in order to obtain this, we modify the algorithm slightly as follows. Let $I : \{1, \dots, n_t\} \rightarrow \{1, \dots, n_t\}$ be the identity operator, i.e. $I(x) = x$. At the beginning of each sweep, i.e. for every $k = q(n_t^2 - n_t)/2$ where $q \in \mathbb{N}$, the SU sets $I_q = I$ and for each $k \in \{q(n_t^2 - n_t)/2 + 1, \dots, (q+1)(n_t^2 - n_t)/2\}$, the SU modifies I_q as follows

$$I_q(l_k) = \begin{cases} l_k & \text{if } a_{l_k l_k} \geq a_{m_k, m_k} \\ m_k & \text{otherwise} \end{cases} \quad (22)$$

At the end of each sweep, i.e. for $k = (q+1)(n_t^2 - n_t)/2$, the SU permutes the columns of \mathbf{W}_k such that \mathbf{W}_k 's l th column becomes its $I_q(l)$'s column. Note that this modification does not require extra transmission cycles since all the additional calculations are carried out at the SU's transmitter. Furthermore, the convergence result in Theorem 3 is still valid. We refer to the OBNSLA after this modification as the modified OBNSLA.

Besides the fact that this modification is necessary for guaranteeing the quadratic convergence rate, as will be shown in the following theorem, it will also be shown that it helps the SU to identify the null space (the last n_r columns of \mathbf{W}_k) blindly without taking extra measurements.

Theorem 4: Let η be the accuracy of the binary search, $\{\lambda_l\}_{l=1}^{n_t}$ be \mathbf{G} 's eigenvalues and let

$$\delta = \min_{\lambda_l \neq \lambda_r} |\lambda_l - \lambda_r|/3 \quad (23)$$

Let P_k be the Frobenius norm of the off diagonal upper triangular part of $\mathbf{A}_k = \mathbf{W}_{k-1}^* \mathbf{G} \mathbf{W}_{k-1}$, where \mathbf{W}_k is defined in (9) and let $m = (n_t^2 - n_t)/2$. Assume that the modified O/BNSLA has reached a stage k , such that $P_k^2 < \delta^2/8$, then

$$P_{k+2m}^2 \leq O\left(\left(\frac{P_{k+m}^2}{\delta}\right)^2\right) + O\left(\frac{\eta P_{k+m}^{3/2}}{\delta}\right) + O\left(\frac{\eta^2 P_{k+m}^{1/2}}{\delta}\right) + 2(n_t^2 - n_t)\eta^2 \|\mathbf{G}\|^2 \quad (24)$$

.

Furthermore, the last $n_t - n_r$ columns of \mathbf{W}_{k+m} inflict minimum interference to the PU; i.e., $\|\mathbf{H}_{ps} \mathbf{w}_i^{k+m}\| \leq \|\mathbf{H}_{ps} \mathbf{w}_j^{k+m}\|$, $\forall 1 \leq j \leq n_r < i \leq n_t$.

Proof: See Appendix B.

Theorem 4 shows that to guarantee the quadratic convergence rate, the accuracy, η , should be much

smaller than P_k^2 ; that is, let k_0 be an integer such that $P_{k_0}^2 < \delta^2/8$, then

$$P_{k_0+2m} \leq O\left(\left(\frac{P_{k_0+m}}{\sqrt{\delta}}\right)^2\right) \quad (25)$$

if $\eta \ll P_{k_0}^2$. This implies that once P_k becomes very small such that $P_k = O(\eta)$, one cannot guarantee that P_{k+2m} will be smaller than P_{k+m}^2 since at $k+1$ it will be $O(\eta)$.

The asymptotic quadratic convergence rate of Theorem 4 is determined by $1/\delta$ where 3δ is the minimal gap between \mathbf{G} 's eigenvalues. In addition, the quadratic convergence rate takes effect only after $P_k^2 < \delta/8$. Such a condition implies that if δ is very small, it will take the modified OBNSLA many cycles to reach its quadratic convergence rate. This is problematic since MIMO wireless channels may have very close singular values (recall that \mathbf{H}_{12} 's square singular values are equal to \mathbf{G} 's first n_r eigenvalues). If we were using the optimal Cyclic Jacobi technique (i.e. no errors because of finite line search accuracy) this would not have practical implications since a quadratic decrease in P_k , which is independent of δ , occurs prior to the phase where $P_k^2 < \delta/8$ [14]. In the following theorem, we extend this result to the modified OBNSLA.

Theorem 5: Let η be the accuracy of the line search, $\{\lambda_l\}_{l=1}^{n_t}$ be \mathbf{G} 's eigenvalues such that there exists a cluster of eigenvalues; i.e., there exists a subset $\{\lambda_{i_l}\}_{l=1}^v \subset \{\lambda_l\}_l^{n_t}$ such that $\lambda_{i_l} = \lambda + \xi_l$, for $l \in L_2 = \{i_1, \dots, i_v\}$, where $\sum_{l=1}^v \xi_l = 0$ and the rest of the non-equal eigenvalues satisfy $\delta_c > 16\sqrt{\sum \xi_l^2}$, where

$$3\delta_c = \min(\Lambda_1 \cup \Lambda_2) \quad (26)$$

$$\begin{aligned} \Lambda_1 &= \{|\lambda_l - \lambda_r| : l \in L \setminus L_2, \lambda_l \neq \lambda_r\} \\ \Lambda_2 &= \{|\lambda_l - \lambda| : l \in L \setminus L_2\} \\ L &= \{1, \dots, n_t\} \end{aligned} \quad (27)$$

Then, once the modified OBNSLA reaches a k such that

$$2\delta_c \sqrt{\sum_{l \in L_2} \xi_l^2} \leq P_k^2 \leq \delta_c^2/8 \quad (28)$$

it satisfies

$$P_{k+2m}^2 \leq O\left(\left(\frac{P_{k+m}}{\delta_c}\right)^4\right) + O\left(\left(\frac{\eta P_{k+m}^3}{\delta_c}\right)\right) + O\left(\left(\frac{\eta^2 P_{k+m}}{\delta_c}\right)\right) + 2(n_t^2 - n_t)\eta^2 \|\mathbf{G}\|^2 \quad (29)$$

Proof: See Appendix C.

Theorem 5 states that in the presence of a single eigenvalue cluster; i.e. $\sqrt{\sum_l \xi_l^2} \ll \delta_c$, and if

$\eta_k = o(P_k)$, the modified O/BNSLA has four convergence regions: The first region is $P_k^2 \geq \delta_c^2/8$, the second is $2\delta_c\sqrt{\sum_l \xi_l^2} \leq P_k^2 \leq \delta_c^2/8$, the third is $\delta/8 \leq P_k^2 \leq 2\delta_c\sqrt{\sum_l \xi_l^2}$ and the fourth is $P_k^2 \leq \min_l \xi_l^2/8$. In the first and the third regions, the modified OBNSLA has at least a linear convergence rate while in the second and fourth regions, it has a quadratic convergence rate. This means that from a practical point of view, a close cluster of eigenvalues; i.e. $\sqrt{\sum_l \xi_l^2}/\delta_c \ll 1$, is not a problem. This is because once the algorithm enters the second convergence region; i.e., it reaches the stage $k = k_2$ such that $2\delta_c\sqrt{\sum_l \xi_l^2} \ll P_{k_2}^2 \leq \delta_c/8$, P_k will decrease quadratically until $k = k_3$ such that $P_{k_3}^2 \leq 2\delta_c\sqrt{\sum_l \xi_l^2}$. But the latter inequality implies that $P_{k_3} \ll P_{k_2}$, a fact that guarantees a significant reduction P_k ; i.e., from P_{k_2} to P_{k_3} with a quadratic rate.

Nevertheless, P_k will eventually decrease quadratically as P_k^2 becomes smaller than $\delta/8$ as required by Theorem 4. This phenomenon is also a characteristic of the Cyclic Jacobi technique [14].

C. The Asymptotic Level of the Interference to the PU

In the previous theorems, we discussed the convergence rate of the sequence P_k ; i.e., the root sum of squares of \mathbf{A}_k 's off diagonal entries. We now consider the maximum level of interference that the SU inflicts on the PU. Our aim here is to relate the asymptotic behavior of the maximum interference to that of P_k , and to obtain bounds on the maximum interference as a function of η . We begin with the following proposition:

Proposition 6: Let \mathbf{T}_k be the SU's pre-coding matrix defined in (20), \mathbf{t}_i^k be its i th column, $Q = \{1, \dots, n_t - n_r\}$, and P_k be the norm of the off diagonal upper triangular (or lower triangular) part of \mathbf{A}_k (where \mathbf{A}_k is defined in (10)). Then

$$\max_{q \in Q} \|\mathbf{H}_{12}\mathbf{t}_q^k\|^2 \leq 2P_k^2 \quad (30)$$

Proof: This is an immediate result of [19, Corollary 6.3.4] which states that for every eigenvalue $\hat{\lambda}$ of $\mathbf{B} + \mathbf{E}$, where \mathbf{B} is an $n_t \times n_t$ Hermitian matrix with eigenvalues $\lambda_i, i = 1, \dots, n_t$, there exists λ_i such that $|\hat{\lambda} - \lambda_i|^2 \leq \|\mathbf{E}\|^2$, where $\|\cdot\|$ is the Forbinus norm. Thus, if one expresses \mathbf{A}_k as $\mathbf{A}_k = \mathbf{B} + \mathbf{E}$, where $\mathbf{B} = \text{diag}(\mathbf{A}_k)$, $\mathbf{E} = \text{offdiag}(\mathbf{A}_k)$, (30) follows. \square

Since the maximum interference to the PU is bounded by $2P_k^2$ (from Proposition 6), it is possible to apply the results of Theorems 4 and 5 and to bound the maximum interference. These bounds are valuable since they relate the asymptotic level of interference to the accuracy of the line search η (which is determined by the SU), thus enabling the SU to control the interference reduction to the PU.

Before obtaining the first bound on the interference, we need the following corollary of Theorem 3:

Corollary 7:

$$\limsup_k P_k^2 \leq \frac{(n_t^2 - n_t)(7 + 2\sqrt{2})\eta^2 \|\mathbf{G}\|^2}{2^{-(n_t-2)(n_t-1)/2}}. \quad (31)$$

Proof: See Appendix D.

From Corollaries 6 and 7 we obtain the following bound:

$$\limsup_k \max_{q \in Q} \|\mathbf{H}_{12} \mathbf{t}_q^k\|^2 \leq \frac{2(n_t^2 - n_t)(7 + 2\sqrt{2})\eta^2 \|\mathbf{G}\|^2}{2^{-(n_t-2)(n_t-1)/2}}, \quad (32)$$

We now derive a tighter bound than (32) which is valid only if the conditions of Theorem 4 are satisfied; i.e., that the OBNSLA is replaced by the modified OBNSLA and that there exists k such that $P_k^2 < \delta^2/8$. In this case, by combining Proposition 6 and Theorem 4, one obtains

$$\max_{q \in Q} \|\mathbf{H}_{12} \mathbf{t}_q^k\|^2 \leq O\left(\frac{P_k^2}{\delta}\right)^2 + O\left(\frac{\eta P_k^{3/2}}{\delta}\right) + O\left(\frac{\eta^2 P_k^{1/2}}{\delta}\right) + 2(n_t^2 - n_t)\eta^2 \|\mathbf{G}\|^2 \quad (33)$$

Furthermore, if P_k becomes sufficiently small such that $\eta > P_k$, the dominant term in the RHS of (33) will be $O(\eta^2)$; i.e., we effectively have:

$$\max_{q \in Q} \|\mathbf{H}_{12} \mathbf{t}_q^k\|^2 \leq 2(n_t^2 - n_t)\eta^2 \|\mathbf{G}\|^2 + O(\eta^{2.5}) \quad (34)$$

V. THE REDUCED COMPLEXITY BLIND NULL SPACE LEARNING (RC-BNSL)

In this section we present the Reduced Complexity MBNSL (RC-MBNSL) algorithm. In the BNSL algorithm, each step requires two binary searches, where each search point is obtained by a transmission cycle. These transmission cycles are the dominant latency factor in the learning process since the rest of the calculations are performed offline at the secondary device processing unit. Roughly speaking (as discussed in Section III-A), the number of rotations (recall that each Jacobi rotation is a learning stage in the M/BNSL algorithms) of the Cyclic Jacobi technique, for daigonalizing a Hermitian matrix \mathbf{G} grows like n_t^2 , the dimension of the matrix \mathbf{G} . In this section it is shown that if \mathbf{G} is low rank, its diagonalization can be simplified to $(n_t - n_r)$ diagonalizations, each one of a single $n_r \times n_r$ matrix. This is possible due to the fact that $\mathbf{G} = \mathbf{H}^* \mathbf{H}$ is an n_r -rank matrix and therefore has an $(n_t - n_r)$ -dimensional null space⁶. The resulting number of stages of each Jacobi sweep grows like $(n_t - n_r)n_r^2$ where in the M/BNSL it grows like n_t^2 . Therefore, the proposed RC-BNSL algorithm is more efficient than the BNSLA if n_t is sufficiently larger than n_r , that is, if the SU has more antenna at its transmitter than the PU has in

⁶If the matrix \mathbf{G} was known, the fact that $\text{rank}(\mathbf{G}) = n_r$ could have been utilized by QR decomposition (which cannot be done blindly) prior to the diagonalization. Note while the SU does not know \mathbf{G} , we assume that he know its rank n_r . This is possible if the SU know what kind of PU uses the channel.

its receiver. This complexity reduction is significant if the SU is a very large MIMO i.e. for wireless communication with tens or even hundreds of antenna [?].

The idea behind the RC-BNSL algorithm is described in the following observation:

Observation 8: Let $\mathbf{H} \in \mathbb{C}^{n_r \times n_t}$ be an n_r -rank ($n_t > n_r$) matrix and let $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_{n_r+1}] \in \mathbb{C}^{n_t \times (n_r+1)}$ be an orthonormal matrix (that is, a matrix whose columns are an orthonormal set i.e. $\mathbf{U}^* \mathbf{U} = \mathbf{I}$) and let $\tilde{\mathbf{H}} = \mathbf{H}\mathbf{U}$. If $\tilde{\mathbf{u}} \in \mathcal{N}(\tilde{\mathbf{H}})$ then $\mathbf{u} = \mathbf{U}\tilde{\mathbf{u}} \in \mathcal{N}(\mathbf{H})$.

Proof: This is due to $\mathbf{H}\mathbf{u} = \mathbf{H}\mathbf{U}\tilde{\mathbf{u}} = \mathbf{0}$ since $\tilde{\mathbf{u}} \in \mathcal{N}(\mathbf{H}\mathbf{U})$.

The RC-BNSL algorithm is carried out as follows: The secondary user begins with an initial pre-coding matrix $\mathbf{U}^{(0)} \in \mathbb{C}^{n_r \times (n_t - n_r)}$ which is composed of the last $n_t - n_r$ columns of some unitary matrix $\mathbf{W} \in \mathbb{C}^{n_t \times n_t}$. Let $\mathbf{H}_{\text{eq}}^{(1)} = \mathbf{H}\mathbf{U}^{(0)} \in \mathbb{C}^{n_r \times (n_r+1)}$, then there exists at least one degree of freedom in this channel. The SU can apply the BNSL algorithm on $\mathbf{G}^{(1)} = \mathbf{H}_{\text{eq}}^{(1)*} \mathbf{H}_{\text{eq}}^{(1)}$ and obtains a pre-coding matrix $\mathbf{U}_k^{(1)}$ such that $\tilde{\mathbf{U}}^{(1)} = \lim_{k \rightarrow \infty} \tilde{\mathbf{U}}_k^{(1)}$ and that

$$\mathbf{\Lambda}^{(1)} = \tilde{\mathbf{U}}^{(1)*} \mathbf{G}^{(1)} \tilde{\mathbf{U}}^{(1)} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & \lambda_{n_r}^{(1)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_1^{(1)} \end{bmatrix} \quad (35)$$

Now that $\mathbf{G}^{(1)}$ is diagonalized, the first degree of freedom is given by $\mathbf{v}^{(1)} = \mathbf{U}^{(0)} \tilde{\mathbf{u}}_1^{(1)}$ where $\tilde{\mathbf{u}}_1^{(1)}$ is the first column of $\tilde{\mathbf{U}}^{(1)}$ (that lies in the null space of $\mathbf{G}^{(1)}$). The SU then can gain an additional degree of freedom by applying the BNSL algorithm on the following $(n_1 + 1)$ equivalent channel

$$\mathbf{H}_{\text{eq}}^{(2)} = \mathbf{H}\mathbf{U}^{(1)}, \quad (36)$$

where $\mathbf{U}^{(1)} \in \mathbb{C}^{n_t \times n_r+1}$ is obtained by concatenating the $n_r - 1$ columns of the initial unitary matrix \mathbf{W} with the last n_r column of $\tilde{\mathbf{U}}^{(1)}$ multiplied by $\mathbf{U}^{(0)}$, i.e. let $\hat{\mathbf{U}}^{(1)} = [\tilde{\mathbf{u}}_2^{(1)}, \dots, \tilde{\mathbf{u}}_{n_r+1}^{(1)}]$, then $\mathbf{U}^{(1)} = [\mathbf{w}_{n_r-1}, \mathbf{U}^{(0)} \hat{\mathbf{U}}^{(1)}]$. This equivalent channel is then diagonalized using the BNSL algorithm to obtain $\tilde{\mathbf{U}}^{(2)}$. We now have two degrees of freedom given by $\mathbf{v}^{(2)} = \mathbf{U}^{(1)} \tilde{\mathbf{u}}_1^{(2)}$ and $\mathbf{v}^{(1)}$. This process is repeated until all \mathbf{W} 's columns are used. The RC-BNSL algorithm is summarized in Algorithm 3.

We conclude with the following corollary, which extends the convergence analysis presented in this section to the BNSLA.

Corollary 9: Theorems 3, 4, 5, Proposition 6 and Corollary 7 apply to the BNSLA presented in [4].

Proof: The proofs of Theorems 3, 4, 5, Proposition 6 and Corollary 7 rely on the fact that the only difference between the CJT and the OBNSLA is in the rotation angles. In the OBNSLA, the CJT's

Algorithm 3 The RCBNSL algorithm

Input: $n_t, n_r, \{\tilde{h}_v\}_{v \in \mathbb{N}}$, defined in (5), $\mathbf{W} \in \mathbb{C}^{n_t \times n_t}$, s.t. $\mathbf{W}^* \mathbf{W} = \mathbf{I}$
Initialize: $\mathbf{U}^{(0)} = [\mathbf{w}_{n_t - n_r}, \dots, \mathbf{w}_{n_t}]$
for $m = 1, \dots, n_t - n_r$ **do**
 $a_n(\mathbf{x}_1, \mathbf{x}_2) \leftarrow \tilde{h}_n(\mathbf{U}^{(m-1)} \mathbf{x})$
 $\tilde{\mathbf{U}}^{(m)} = \text{MBNSL}(\{a_v\}_{v \in \mathbb{N}}, n_r + 1, n)$
 $\mathbf{v}_m = \mathbf{U}^{(m-1)} \tilde{\mathbf{u}}_1^{(m)}$
 $\check{\mathbf{U}}_m = [\tilde{\mathbf{u}}_2^{(m)}, \dots, \tilde{\mathbf{u}}_{n_r+1}^{(m)}]$
 $\mathbf{U}^{(m)} = [\mathbf{w}_{n_t - n_r - m}^{(m)}, \mathbf{U}^{(m-1)} \check{\mathbf{U}}^{(m)}]$
end for

rotation angles, θ_k^J, ϕ_k^J , are approximated according to (17). Furthermore, note that the BNSLA and the OBNSLA are identical except for the way in which each algorithm determines its SMI (which are not identical SMIs) before invoking the binary search. However, (17) is satisfied as long as each SMI contains the desired minimum point. Because the latter is satisfied by both algorithms, as indicated by Proposition 2 for the OBNSLA and by Proposition 3 in [4] for the BNSLA, Theorems 3, 4, 5, Proposition 6 and Corollary 7 apply to the BNSLA. \square

VI. SIMULATIONS

A. Non-Asymptotic Comparison of the BNSLA and the OBNSLA

In this section we compare the OBNSLA to the BNSLA. In this simulation the PU performs a power adaptation every 1 msec to maintain a target 10 dB SINR at the receiver, and the SU inflicts interference on the PU and measures $q(n)$ by listening to the PU signal's power at the SU-Rx⁷. Fig. 4. presents the interference reduction of the BNSLA and the OBNSLA as a function of \mathbf{H}_{ps} 's Doppler spread. The result shows that both algorithms have similar performances.

An important practical issue in the implementation the OBNSLA and the BNSLA is granularity in the PU's SINR; i.e., let $\text{SINR}(n)$ be the PU's SINR at the n th TC, and let $\text{SINR}_q = [\text{SINR}_q^{\min}, \text{SINR}_q^{\max})$, $q = 1, \dots, Q$, be the granularity of the PU's SINR, that is, if $\text{SINR}(n), \text{SINR}(n+1) \in \text{SINR}_q$, the PU will not modify its transmission scheme. Note that if $\|\mathbf{H}\tilde{\mathbf{x}}(n)\|^2 < \|\mathbf{H}\tilde{\mathbf{x}}(n-1)\|^2$ and the difference between them is small, such that $\text{SINR}(n), \text{SINR}(n-1) \in \text{SINR}_q$, the SU will observe $\tilde{h}(\tilde{\mathbf{x}}(n), \tilde{\mathbf{x}}(n-1)) = 1$, which falsely indicates that $\|\mathbf{H}\tilde{\mathbf{x}}(n)\|^2 \geq \|\mathbf{H}\tilde{\mathbf{x}}(n-1)\|^2$. Such errors may be significant since the binary search is directed to the wrong interval. A full theoretical convergence

⁷The SU sets $q(n) = \frac{1}{N} \sum_{t=Nn}^{Nn+N'-1} \|\mathbf{y}_s(t) - \bar{\mathbf{y}}_s\|^2$, where $\bar{\mathbf{y}}_s$ is the average of $\mathbf{y}_s(t)$ over $t = N(n-1) + N', \dots, Nn + N' - 1$. The consideration for choosing such q are described in [4, Sec II-B1].

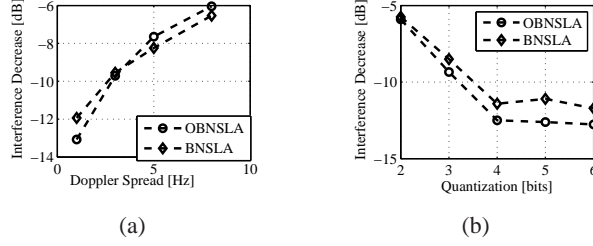


Fig. 4. Interference reduction after a single Jacobi sweep as a function of \mathbf{H}_{ps} 's Doppler spread. The maximum power constraint of each Tx is 23 dBm. The channels' path-losses are calculated as $128.1 + 37.6 \log_{10}(R)$, where R is the distance between the Rx and the TX in meters as used by the 3GPP (see page 61 3GPP Technical Report 36.814). The locations of the PU-TX is randomly chosen from a uniform distribution over a 300 m disk, and the locations of the SU-Tx and the SU-Rx are randomly chosen from a uniform distribution over a 400 m disk. Both disks are centered at the location of the PU-Rx. The minimum distance between the PU-Rx to the PU-Tx, and the PU-Rx to the SU-Tx is 20 m and 100 m, respectively. For each t the entries of the channel matrices $\mathbf{H}_{pp}(t)$, $\mathbf{H}_{sp}(t)$, $\mathbf{H}_{ss}(t)$ are i.i.d. where each entry is 15 KHz flat fading Rayleigh channel with 15 Hz Doppler spread in. $\mathbf{H}_{ps}(t)$ has distribution as the other channel, except for its Doppler spread which is given by the horizontal axes of both subfigures. All channels are generated using the Improved Rayleigh Fading Channel Simulator [20]. The noise level at the receivers is -121 dBm and the SU transmit power during the learning process is 5 dBm. The numbers of antennas are $n_{ts} = 2$, $n_{tp} = 2$, $n_{rp} = 1$, $n_{rs} = 2$.

analysis of this problem is an important topic for future research. In this paper, we test this problem using a simulation. Fig. 4(b) presents simulation results for a scenario where the PU's power control process is based on a quantized measurement of the PU's SINR in the range -5 dB to 20 dB. It is shown that the interference reduction is not improved for more than 4 bit quantization. This means that small granularity does have a practical affect on the performance of the O/BNSLA.

B. Asymptotic analysis

We now compare the asymptotic properties of the OBNSLA to the bounds derived in Section IV. Figure 5 presents simulation results for the OBNSLA under optimal conditions; i.e., $q(n)$ is perfectly observed, for different levels of line search accuracies. Figure 5(a) depicts P_k and the bound on it as given in (32), versus complete OBNSLA sweeps; i.e., $(n_t^2 - n_t)/2$ learning phases. It shows that for sufficiently small η the OBNSLA converges quadratically. The quadratic decrease breaks down when the value of P_k becomes as small as an order of magnitude of η . This result is consistent with Theorem 4. Figure 5(b) depicts the interference decrease and the bound on it as given in (34) versus the number of transmission cycles. It shows that the asymptotic level of the interference to the PU is bounded by $O(\eta^2)$.

The bounds in this paper are derived under the assumption that the OC holds perfectly. In practice, however, $q(n)$ is affected by measurement error such as noise. Furthermore, the channel matrices \mathbf{H}_{pp} , \mathbf{H}_{sp} vary with time, a fact that may also affect the function $q(n)$. For example, consider a case where \mathbf{H}_{pp} is

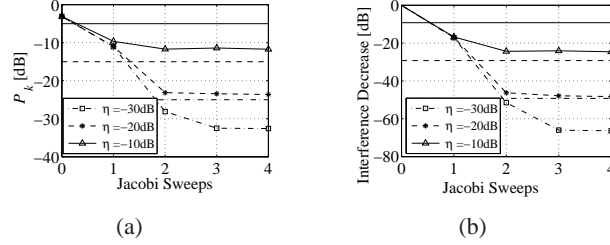


Fig. 5. Simulation results for different values η of the OBNSL algorithm to obtain the null space of \mathbf{H} with $n_t = 3$ transmitting antennas and $n_r = 2$ antennas at the PU receiver. The matrix $\mathbf{G} = \mathbf{H}^* \mathbf{H}$ was normalized such that $\|\mathbf{G}\|^2 = 1$. The unmarked lines in (a) and (b) represent the asymptotic upper bound of (32) and (34) respectively on the corresponding marked line; e.g. the solid unmarked line is a bound on the solid line with squares. We used 200 Monte-Carlo trials where the entries of \mathbf{H} are i.i.d. complex Gaussian random variables.

a scalar and the SU extracts $q(n)$ by listening to variations in the PU's transmit power. Assume further that the gain of \mathbf{H}_{pp} is decreased between the two consecutive transmission cycles and that the PU compensates for this by increasing its transmit power. In this case the SU might mistakenly deduce that the interference it inflicts on the PU has increased. In addition, \mathbf{H}_{ps} is also time-varying, which leads to some discrepancy between the estimated null space and the true null space. In what follows, we show in simulations that the derived bounds are still useful in practice if the values of η are not “too”. A full theoretical convergence analysis of the O/BNSLA in practical conditions, which extends the bounds derived in this paper to account for measurement noise and time variations in the channel is a topic for future research, beyond the scope of this paper.

Figure 6 presents simulation results for an identical scenario as in Figure 4 except for \mathbf{H}_{ps} which is generated assuming that both the SU-Tx and PU-Rx are fixed. \mathbf{H}_{ps} was generated according to [21], which represents a fixed Rx-Tx channel where one antenna is 1.75 m in height and the second antenna is 25 m in height. The Rician factor and the Doppler spread of \mathbf{H}_{ps} were determined according to Equations (13) and (14) in [21]. The result shows that for an interference reduction smaller or equal to 37 dB, the bound in (34) predicts the behavior of the interference reduction (i.e., it decreases as η^2) of both the OBNSLA and the BNSLA. Furthermore, it is shown that the BNSLA and the OBNSLA have the same asymptotic properties; i.e. convergence rate and asymptotic interference reduction.

VII. SUMMARY AND FUTURE RESEARCH

This paper proposed a new algorithm termed the One-bit Null Space Learning Algorithm (OBNSLA), which enables a MIMO CR SU to learn the null space of the interference channel to the PU by observing a binary function that indicates the variations (increase or decrease) in the PU's SINR. Such information can be extracted, for instance, from the quantized version of the variation in the PU's SINR, or in the

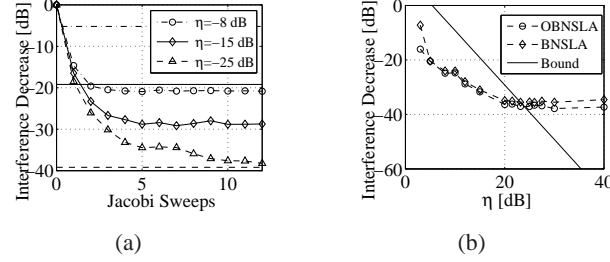


Fig. 6. Interference reduction (marked lines) of the OBNSLA, as a function of Jacobi Sweeps (a), and as a function of η (b). The unmarked lines represent the bound in (34) for the corresponding marked lines with the same pattern, e.g., in Subfigure (a) the dotted-dashed unmarked line represent the bound on the OBNSLA's interference reduction with $\eta = -8$ dB, which is represented by the dotted-dashed line that is marked with circles. The numbers of antennas are $n_{t_s} = 2$, $n_{t_p} = 2$, $n_{r_p} = 1$, $n_{r_s} = 2$. The results were averaged over 1000 Monte-Carlo trials.

PU's modulation. We also provided a convergence analysis of the OBNSLA, which also applies to the Blind Null Space Learning Algorithm (BNSLA) that was recently proposed [4]. It was shown that the two algorithms maintain the “good” convergence properties of the Cyclic Jacobi technique, namely a global linear and an asymptotically quadratic convergence rate. It was also shown in simulations that just like in the Cyclic Jacobi technique, the OBNSLA and the BNSLA reach their quadratic convergence rates in only three to four cycles. In addition, we derived asymptotic bounds on the maximum level of interference that the SU inflicts on the PU. The derived bounds have important practical implications. Due to the fact that these bounds are functions of a parameter determined by the SU, it enables the SU to control the maximum level of interference caused to the PU. This gives the OBNSLA (or the BNSLA) a useful stopping criterion which guarantees the protection of the PU. The analytical convergence rates and interference bounds were validated by extensive simulations.

We consider the theoretical analysis of the OBNSLA and BNSLA under measurement noise as an important topic for future research. Note that in the presence of noise, the analysis of the two algorithms is not identical since the BNSLA relies on a continuous-valued function of the PU's SINR, whereas the OBNSLA relies on a binary function. Noise, which is continuous-valued, will thus affect these functions and hence the performance and convergence of the two algorithms quite differently.

APPENDIX A

PROOF OF THEOREM 3

Consider the first sweep of the BNSL algorithm; i.e. $k = 1, 2, \dots, n_t(n_t - 1)/2$. Denote the number of rotated elements in the l th row by $b_l = n_t - l$ and let

$$c_l = \sum_{j=1}^l b_j = (2n_t - 1 - l)l/2; \quad Z(l, k) = \sum_{j=1}^{n_t-l} |[\mathbf{A}_k]_{j+l,l}|^2; \quad W(l, k) = \sum_{j=l+1}^{n_t-1} Z(j, k) \quad (37)$$

Note that $W(0, k) = P_k^2$. In every sweep, each entry is eliminated once; we therefore denote \mathbf{A}_k 's p, q entry before its annihilation as $g_{q,p}(t)$ where t denotes the number of changes since $k = 0$. After $g_{q,p}(t)$ is annihilated once, it will be denoted by $\tilde{g}_{q,p}(\tilde{t})$ where \tilde{t} is the number of changes after the annihilation. The diagonal entries of \mathbf{A}_k will be denoted by x since we are not interested in their values in the course of the proof. This is illustrated in the following example of a 4×4 matrix

$$\begin{aligned} \mathbf{A}_0 = \mathbf{G} & \quad \mathbf{A}_1 \\ \begin{pmatrix} g_{1,1}(0) & g_{1,2}(0) & g_{1,3}(0) & g_{1,4}(0) \\ g_{2,1}(0) & g_{2,2}(0) & g_{2,3}(0) & g_{2,4}(0) \\ g_{3,1}(0) & g_{3,2}(0) & g_{3,3}(0) & g_{3,4}(0) \\ g_{4,1}(0) & g_{4,2}(0) & g_{4,3}(0) & g_{4,4}(0) \end{pmatrix} & \quad \begin{pmatrix} x & \epsilon & g_{1,3}(1) & g_{1,4}(1) \\ \epsilon & x & g_{2,3}(1) & g_{2,4}(1) \\ g_{3,1}(1) & g_{3,2}(1) & x & g_{3,4}(0) \\ g_{4,1}(1) & g_{4,2}(1) & g_{4,3}(0) & x \end{pmatrix} \\ \mathbf{A}_2 & \quad \mathbf{A}_3 \\ \begin{pmatrix} x & \tilde{g}_{1,2}(0) & \epsilon & g_{1,4}(2) \\ \tilde{g}_{2,1}(0) & x & g_{2,3}(2) & g_{2,4}(1) \\ \epsilon & g_{3,2}(2) & x & g_{3,4}(1) \\ g_{4,1}(2) & g_{4,2}(1) & g_{4,3}(1) & x \end{pmatrix} & \quad \begin{pmatrix} x & \tilde{g}_{1,2}(1) & \tilde{g}_{1,3}(0) & \epsilon \\ \tilde{g}_{2,1}(1) & x & g_{2,3}(2) & g_{2,4}(2) \\ \tilde{g}_{3,1}(0) & g_{3,2}(2) & x & g_{3,4}(2) \\ \epsilon & g_{4,2}(2) & g_{4,3}(2) & x \end{pmatrix} \end{aligned} \quad (38)$$

For arbitrary n_t , after the first c_1 sweeps \mathbf{A}_{c_1} 's first column is equal to the following vector:

$$[x, \tilde{g}_{2,1}(n_t - 3), \dots, \tilde{g}_{n_t-1,1}(0), \epsilon_{c_1}]^T \quad (39)$$

and

$$Z(1, c_1) \leq |\tilde{g}_{2,1}(n_t - 3)|^2 + \dots + |\tilde{g}_{n_t-1,1}(0)|^2 + |\epsilon_{c_1}|^2 \quad (40)$$

From (10) it follows that for $q = 2, \dots, n_t$,

$$\begin{aligned}
\tilde{g}_{q,1}(n_t - q - 1) &= \cos(\theta_{n_t-1}) \tilde{g}_{q,1}(n - q - 2) - e^{i\phi_{n_t-1}} g_{q,n_t}(1) \sin(\theta_{n_t-1}) \\
&\vdots \\
\tilde{g}_{q,1}(1) &= \cos(\theta_{q+1}) \tilde{g}_{q,1}(0) - e^{i\phi_3} g_{q,q+2}(1) \sin(\theta_3) \\
\tilde{g}_{q,1}(0) &= \epsilon_{q-1} \cos(\theta_q) - e^{i\phi_q} g_{q,q+1}(1) \sin(\theta_q)
\end{aligned} \tag{41}$$

where $\tilde{g}_{q,1}(-1) = \epsilon_1$. The following bounds on $\{\tilde{g}_{q,1}(l)\}_{l=0}^{n_t-q-1}$ are obtained recursively (i.e., by obtaining a bound on $\tilde{g}_{q,1}(0)$, substituting and obtaining a bound on $\tilde{g}_{q,1}(1)$ and so on)

$$\begin{aligned}
\tilde{g}_{q,1}(n_t - q - 1) &\leq |\epsilon_{q-1} \prod_{v=q}^{n_t-1} \cos(\theta_v) - \sum_{j=q}^{n_t-1} e^{i\phi_j} \sin(\theta_j) g_{q,j+1}(1) \prod_{v=j+1}^{n_t-1} \cos(\theta_v)| \\
&\leq |\mathbf{v}(q)^T \mathbf{y}(q)| + \epsilon \prod_{v=q}^{n_t-1} \cos(\theta_v)
\end{aligned} \tag{42}$$

where $\mathbf{v}, \mathbf{y} \in \mathbb{C}^{n_t-q}$ such that $[\mathbf{v}(q)]_j = e^{i\phi_{j+q-1}} g_{q,j+q}(1)$, $[\mathbf{y}(q)]_j = \sin(\theta_{j+q-1}) \prod_{v=j+q}^{n_t-1} \cos(\theta_v)$, $j = 1, \dots, n_t - q$, and $\epsilon = \max_q |\epsilon_q|$. It follows that

$$|\tilde{g}_{q,1}(n_t - q - 1)|^2 \leq |\mathbf{y}^T(q) \mathbf{v}(q)|^2 + |\epsilon|^2 \prod_{v=q}^{n_t-1} \cos^2(\theta_v) \leq \|\mathbf{y}(q)\|^2 \|\mathbf{v}(q)\|^2 + |\epsilon|^2 \prod_{v=q}^{n_t-1} \cos^2(\theta_v) \tag{43}$$

Proposition 10:

$$\|\mathbf{y}(q)\|^2 = 1 - \prod_{i=q}^{n-1} \cos(\theta_i) \tag{44}$$

Proof: This is shown by induction. By definition

$$\|\mathbf{y}(q)\|^2 = \sum_{i=q}^{n-1} \sin^2(\theta_i) \prod_{v=i+1}^{n-1} \cos^2(\theta_v) \tag{45}$$

where

$$\prod_{i=l}^m v_i \triangleq 1, \text{ if } l > m \tag{46}$$

Assume that (44) is true for $n = m \in \mathbb{N}$, then, for $m + 1$ (44) and (45) yields

$$\begin{aligned}
\sum_{i=q}^m \sin^2(\theta_i) \prod_{v=i+1}^m \cos^2(\theta_v) &= \sum_{i=q}^{m-1} \sin^2(\theta_i) \prod_{v=i+1}^m \cos^2(\theta_v) + \sin^2(\theta_m) \prod_{v=m+1}^m \cos^2(\theta_v) \\
&= \cos^2(\theta_m) \sum_{i=q}^{m-1} \sin^2(\theta_i) \prod_{v=i+1}^{m-1} \cos^2(\theta_v) + \sin^2(\theta_m),
\end{aligned} \tag{47}$$

where the last equality is due to (46). According to the supposition (44)

$$\cos^2(\theta_m) \left(1 - \prod_{i=q}^{m-1} \cos^2(\theta_i)\right) + \sin^2(\theta_m) = 1 - \prod_{i=q}^m \cos^2(\theta_i), \tag{48}$$

which establishes the desired result. \square

By substituting Proposition 10 into (43) one obtains

$$|\tilde{g}_{q,1}(n_t - q - 1)|^2 \leq \underbrace{\left(\sum_{i=1}^{n_t-q} |g_{q,i+q}(1)|^2 \right)}_{=Z(q,c_1)} \left(1 - \prod_{i=c_0+q}^{n_t-1} \cos^2(\theta_i) \right) + |\epsilon|^2 \underbrace{\prod_{v=q}^{n_t-1} \cos^2(\theta_v)}_{\leq 1} \quad (49)$$

thus,

$$|\tilde{g}_{q,1}(n_t - q - 1)|^2 \leq \left(1 - \prod_{i=c_0+q}^{n_t-1} \cos^2(\theta_i) \right) Z(q, c_1) + |\epsilon|^2 \quad (50)$$

and by summing both sides of (50) over $q = 2, \dots, n_t$

$$Z(1, c_1) \leq \sum_{q=2}^{n_t} \left(1 - \prod_{i=c_0+q}^{n_t-1} \cos^2(\theta_i) \right) Z(q, c_1) + (n_t - 1)|\epsilon|^2 \quad (51)$$

$$\leq \left(1 - \prod_{i=c_0+2}^{c_1} \cos^2(\theta_i) \right) \underbrace{\sum_{q=2}^n Z(q, c_1)}_{W(1, c_1)} + (n_t - 1)|\epsilon|^2 \quad (52)$$

$$\leq \left(1 - \prod_{i=c_0+2}^{c_1} \cos^2(\theta_i) \right) W(0, 0) + (n_t - 1)|\epsilon|^2 \quad (53)$$

where the last inequality is due to $P_{c_1} = W(1, c_1) + Z(1, c_1)$, $W(0, 0) = P_0$, and because P_k is a monotonically decreasing sequence⁸. It follows that

$$Z(1, c_1) = \sin^2(\Psi_{c_0+2, c_1}) W(0, 0) + (n_t - 1)|\epsilon|^2 \quad (54)$$

where

$$\sin^2(\Psi_{c_{l-1}+2, c_l}) = 1 - \prod_{i=c_{l-1}+2}^{c_l} \cos^2(\tilde{\theta}_i) \quad (55)$$

and $\tilde{\theta}_i$ is an angle that satisfies $|\tilde{\theta}_i| \leq |\theta_i|$. Thus,

$$P_{c_1} = W(1, c_1) + Z(1, c_1) \leq W(0, 0) = P_0 \quad (56)$$

substituting (54) we obtain

$$W(1, c_1) \leq W(0, 0) \cos^2(\Psi_{2, c_1}) - (n_t - 1)|\epsilon|^2 \quad (57)$$

⁸Forsythe and Henrici [9] showed that the sequence P_k is a monotonically decreasing sequence.

Now that this relation is established, it can be applied to \mathbf{A}_{c_l} 's lower $(n_t - l) \times (n_t - l)$ block-diagonal, thus

$$W(l, c_l) \leq W(l-1, c_{l-1}) \cos^2(\Psi_{c_{l-1}+2, c_l}) - (n_t - l)|\epsilon|^2 \quad (58)$$

By substituting (58) recursively into itself, one obtains

$$W(l, c_l) \leq W(0, 0) \prod_{j=1}^l \cos^2(\Psi_{c_{j-1}+2, c_j}) - \epsilon^2 \sum_{j=1}^l b_j \prod_{v=j+1}^l \cos^2(\Psi_{c_{v-1}+2, c_v}) \quad (59)$$

Thus

$$\begin{aligned} Z(l, c_l) &= \sin^2(\Psi_{c_{l-1}+2, c_l}) W(l-1, c_{l-1}) + (n_t - l)|\epsilon|^2 \leq W(0, 0) \sin^2(\Psi_{c_{l-1}+2, c_l}) \prod_{j=1}^{l-1} \cos^2(\Psi_{c_{j-1}+2, c_j}) \\ &\quad - |\epsilon|^2 \sum_{j=1}^{l-1} b_j \prod_{v=j+1}^{l-1} \cos^2(\Psi_{c_{v-1}+2, c_v}) + (n_t - l)|\epsilon|^2 \end{aligned} \quad (60)$$

After a complete sweep

$$\begin{aligned} P_{c_{n_t-1}}^2 &= \sum_{l=1}^{n_t-2} Z(l, c_{n_t-1}) + |\epsilon|^2 = \sum_{l=1}^{n_t-2} Z(l, c_l) \leq W(0, 0) \sum_{l=1}^{n_t-2} \sin^2(\Psi_{c_{l-1}+2, c_l}) \prod_{j=1}^{l-1} \cos^2(\Psi_{c_{j-1}+2, c_j}) \\ &\quad - \sum_{l=1}^{n_t-2} |\epsilon|^2 \sum_{j=1}^{l-1} b_j \prod_{v=j+1}^{l-1} \cos^2(\Psi_{c_{v-1}+2, c_v}) + |\epsilon|^2 \sum_{l=1}^{n_t-1} (n_t - l) \end{aligned} \quad (61)$$

where the first equality is due to the fact that for $k = c_l + 1, \dots, c_{n_t}$, the sum of squares of the l th column remains unchanged; thus, $Z(l, k) = Z(l, c_l), \forall k > c_l$. Similar to proposition 10, it can be shown that $\sum_{l=1}^n \sin^2(\tau_l) \prod_{j=1}^{l-1} \cos^2(\tau_j) = 1 - \prod_{j=1}^n \cos^2(\tau_j)$. Thus

$$P_{c_{n-1}}^2 \leq W(0, 0) \left(1 - \prod_{j=1}^{n-2} \cos^2(\Psi_{c_{j-1}+2, c_j}) \right) - \sum_{l=1}^{n-2} \epsilon^2 \sum_{j=1}^{l-1} b_j \prod_{v=j+1}^{l-1} \cos^2(\Psi_{c_{v-1}+2, c_v}) + |\epsilon|^2 \sum_{l=1}^{n-1} b_l \quad (62)$$

From (55) we have

$$\cos^2(\Psi_{c_{l-1}+2, c_l}) \geq \prod_{v=c_{l-1}+2}^{c_{k_l}} \cos^2(\theta_v) \quad (63)$$

and therefore

$$\begin{aligned} P_{c_{n-1}}^2 &\leq W(0, 0) \left(1 - \prod_{j=1}^{n-2} \prod_{v=c_{j-1}+2}^{c_j} \cos^2(\theta_v) \right) \\ &\quad - \sum_{l=1}^{n-2} |\epsilon|^2 \sum_{j=1}^{l-1} (n-j) \prod_{v=j+1}^{l-1} \prod_{r=c_{v-1}+2}^{c_v} \cos^2(\theta_r) + \frac{|\epsilon|^2(n^2-n)}{2} \end{aligned} \quad (64)$$

Recall that $|\theta_i| < \pi/4$, therefore

$$\begin{aligned} P_{c_{n-1}}^2 &\leq W(0, 0) \left(1 - 2^{-(n-2)(n-1)/2} \right) - |\epsilon|^2 \left(\sum_{l=1}^{n-2} \sum_{j=1}^{l-1} (n-j) 2^{\frac{l^2}{2} - ln + \frac{l}{2} + 9n - 45} - \frac{(n^2-n)}{2} \right) \\ &\leq W(0) \left(1 - 2^{-(n-2)(n-1)/2} \right) + |\epsilon|^2 \frac{(n^2-n)}{2} \end{aligned} \quad (65)$$

It remains to relate ϵ to the accuracy of the line search η . Note that the error ϵ in (65) is due to (17) which is a result of the two finite-accuracy (of η accuracy) line-searches in (12), and (13). If η were zero, \mathbf{A}_k 's l, m off diagonal entry would be zero after the k th sweep, i.e.

$$u(\theta_k^J, \phi_k^J) = 0 \quad (66)$$

where

$$u(\theta, \phi) \triangleq |[\mathbf{R}_{l,m}(\theta, \phi) \mathbf{A}_k \mathbf{R}_{l,m}^*(\theta, \phi)]_{l,m}|^2 = u_1(\theta, \phi) + u_2(\theta, \phi), \quad (67)$$

$$\begin{aligned} u_1(\theta, \phi) &= 4(a_{l,m}^k)^2 \sin^2(\gamma_{l,m} + \phi) \\ u_2(\theta, \phi) &= \left(2 \cos(2\theta) a_{l,m}^k \cos(\angle a_{l,m}^k + \phi) + \sin(2\theta) (a_{l,l}^k - a_{m,m}^k) \right)^2 \end{aligned} \quad (68)$$

and (θ_k^J, ϕ_k^J) is the value given in Theorem 1 when substituting $\mathbf{G} = \mathbf{A}_k$. Recall that $(\hat{\theta}_k^J, \hat{\phi}_k^J)$ (see (17)) is the non optimal value that is obtained by the two line searches, then

$$|\epsilon|^2 = \max_k u(\hat{\theta}_k^J, \hat{\phi}_k^J) \quad (69)$$

The error $u(\hat{\theta}_k^J, \hat{\phi}_k^J)$ can be bounded because $\phi_k^J = \angle a_{l,m}^k$, thus $\hat{\phi}_k^J = -\angle a_{l,m}^k + \eta_\phi$ where $|\eta_\phi| < \eta$, and

$$u_1(\hat{\theta}_k^J, \hat{\phi}_k^J) = 4(a_{l,m}^k)^2 \sin^2(\angle a_{l,m}^k + \hat{\phi}_k^J) \leq 4(a_{l,m}^k)^2 \eta^2 \leq 2\|\mathbf{G}\| \eta^2 \quad (70)$$

$$u_2(\hat{\theta}_k^J, \hat{\phi}_k^J) = \left(2a_{l,m}^k \cos(\eta_\phi) \cos^2(2\hat{\theta}_k^J) + \sin(2\hat{\theta}_k^J) (a_{l,l}^k - a_{m,m}^k) \right)^2 \quad (71)$$

To bound $u_2(\hat{\theta}_k^J, \hat{\phi}_k^J)$, note that if $a_{ll}^k = a_{mm}^k$, then $\hat{\theta}_k^J = \theta_k^J \in \{0, \pi/4\}$ since the line search will not miss these points. Now for the case where $a_{ll}^k \neq a_{mm}^k$ we have $\hat{\theta}_k^J = \theta_k^s + \eta_\theta$ where

$$\theta_k^s = \frac{1}{2} \tan^{-1}(x_k) \quad (72)$$

and

$$x_k = \frac{2|a_{l,m}^k| \cos(\eta_\phi)}{a_{m,m}^k - a_{l,l}^k} \quad (73)$$

Note that

$$u_2(\hat{\theta}_k^J, \hat{\phi}_k^J) = \left(2 \cos(\eta_\phi) a_{l,m}^k (\cos(2\theta_k^s) - 2\eta_\theta \sin(2\theta_k^s)) + (a_{l,l}^k - a_{m,m}^k) (\sin(2\theta_k^s) + 2 \cos(2\theta_k^s) \eta_\theta) \right)^2$$

where (θ^*, ϕ^*) is a point on the line that connects the points (θ_k^J, ϕ_k^J) , $(\hat{\theta}_k^J, \hat{\phi}_k^J)$. By substituting (72) we obtain

$$u_2(\hat{\theta}_k^J, \hat{\phi}_k^J) = \left(\frac{2 \cos(\eta_\phi) a_{l,m}^k + x_k a_{l,l}^k - x_k a_{m,m}^k}{\sqrt{x_k^2 + 1}} - 4\eta_\theta \sin(2\theta^*) \cos(\eta_\phi) a_{l,m}^k + 2\eta_\theta \cos(2\theta^*) (a_{l,l}^k - a_{m,m}^k) \right)^2 \quad (74)$$

Using (73) and the fact that the sinusoidal is bounded by one, and because $|\eta_\theta| \leq \eta$, it follows that

$$\begin{aligned} u_2(\hat{\theta}_k^J, \hat{\phi}_k^J) &\leq 4\eta^2 \left(2|\sin(2\theta^*)| a_{l,m}^k + \cos(2\theta^*) |a_{l,l}^k - a_{m,m}^k| \right)^2 \\ &\leq 4\eta^2 \left(4\sin^2(2\theta^*) |a_{l,m}^k|^2 + 2\sin(4\theta^*) |a_{l,m}^k| |a_{l,l}^k - a_{m,m}^k| + \cos^2(2\theta^*) |a_{l,l}^k - a_{m,m}^k|^2 \right) \\ &\leq 4\eta^2 \left(2|a_{l,m}^k|^2 + 2\sin(4\theta^*) |a_{l,m}^k| |a_{l,l}^k - a_{m,m}^k| + |a_{l,l}^k - a_{m,m}^k|^2 + 2|a_{l,m}^k|^2 \right) \end{aligned} \quad (75)$$

$$u_2(\hat{\theta}_k^J, \hat{\phi}_k^J) \leq 4\eta^2 (2\|\mathbf{G}\|^2 + \sqrt{2}\|\mathbf{G}\|\|\mathbf{G}\| + \|\mathbf{G}\|^2) \quad (76)$$

Thus

$$|\epsilon|^2 = \max_k u(\hat{\theta}_k^J, \hat{\phi}_k^J) \leq 2(7 + 2\sqrt{2})\eta^2 \|\mathbf{G}\|^2 \quad (77)$$

This expression is substituted into (65) and the desired result follows. \square

APPENDIX B

PROOF OF THEOREM 4

Without loss of generality, we assume that $W(0, 0) \leq \delta^2/8$ where $W(k, l)$ is defined in (37)⁹. We first prove the theorem assuming that \mathbf{G} 's eigenvalues are all distinct. From (42) it follows that

$$|\tilde{g}_{q,1}(n_t - q - 1)|^2 \leq \sum_{j=q}^{n_t-1} \sin^2(\theta_j) |g_{q,j+1}(1)|^2 + \epsilon^2 \prod_{v=q}^{n_t-1} \cos^2(\theta_v) \quad (78)$$

Similar to the derivation of (50), but without applying Proposition 10, one obtains

$$|\tilde{g}_{q,1}(n_t - q - 1)|^2 \leq Z(q, c_1) \sum_{j=q}^{n_t-1} \sin^2(\theta_j) + |\epsilon|^2 \leq Z(q, c_1) \sum_{j=2}^{n_t-1} \sin^2(\theta_j) + |\epsilon|^2 \quad (79)$$

and by summing both sides of (79) (similar to the derivation of (51)) over $q = 2, \dots, n_t$ it follows that

$$Z(1, c_1) \leq \left(\sum_{j=2}^{n_t-1} \sin^2(\theta_j) \right) \underbrace{\sum_{q=2}^{n_t} Z(q, c_1)}_{W(1, c_1)} + (n_t - 1)|\epsilon|^2 \leq \left(\sum_{j=2}^{n_t-1} \sin^2(\theta_j) \right) W(0, 0) + (n_t - 1)|\epsilon|^2$$

Now that we have established this relation we can apply it to the reduced $n_t - l + 1$ lower block

⁹ I.e. let k_0 be the smallest integer such that $P_{k_0} < \delta^2/8$ and $(l_{k_0}, m_{k_0}) = (1, 2)$, we set $\mathbf{A}_{k_0} = \mathbf{G}$.

diagonal matrix and obtain $Z(l, c_l) \leq \left(\sum_{j=c_{l-1}+1}^{c_l} \sin^2(\theta_j) \right) W(0, 0) + (n_t - l)|\epsilon|^2$. After a complete sweep we have

$$\begin{aligned} P_{c_{n_t-1}}^2 &\leq \sum_{l=1}^{n_t-2} Z(l, c_{n_t-1}) + |\epsilon|^2 = \sum_{l=1}^{n_t-2} Z(l, c_l) + |\epsilon|^2 \\ &\leq W(0, 0) \sum_{j=1}^{n_t(n_t-1)/2} \sin^2(\theta_j) + |\epsilon|^2 \sum_{l=1}^{n_t-1} (n_t - l) \end{aligned} \quad (80)$$

We now relate $\sum_{j=1}^{n_t(n_t-1)/2} \sin^2(\theta_j)$ to $W(0, 0)$ (recall that $P_0^2 = W(0, 0)$). Note that $|a_{ll}^k - a_{mm}^k|^2 = |a_{ll}^k - \lambda_l - a_{mm}^k + \lambda_m + \lambda_l - \lambda_m|^2 \geq |\lambda_l - \lambda_m|^2 - |a_{ll}^k - \lambda_l|^2 - |a_{mm}^k - \lambda_m|^2$, furthermore, by [12, Theorem 1], there exists a permutation to $\{\lambda_i\}_{i=1}^{n_t}$ such that

$$|a_{ii}^k - \lambda_i| \leq \sqrt{2} P_k, \quad (81)$$

thus,

$$|a_{ii}^k - \lambda_i| \leq \delta/2, \quad (82)$$

and

$$|a_{ll}^k - a_{mm}^k| \geq 2\delta - \delta/2 - \delta/2 = \delta. \quad (83)$$

Recall that the optimal rotation angle satisfies $\tan(2\theta_k^J) = 2|a_{l_k m_k}^k|/|a_{l_k l_k}^k - a_{m_k m_k}^k|$ while the actual the rotation angel is

$$\hat{\theta}_k^J = \theta_k^J + \eta_\theta \quad (84)$$

It follows that

$$\begin{aligned} |\sin^2(\hat{\theta}_k^J)| &\leq |\sin^2(\theta_k^J)| + |\eta_\theta \sin(2\theta_k^J)| \leq \frac{1}{4}|2\theta_k^J|^2 + |\eta_\theta| \tan(2\theta_k^J) \leq \frac{1}{2^2} \tan^2(2\theta_k^J) + |\eta_\theta| \tan(2\theta_k^J) \\ &\leq \frac{|a_{l_k m_k}^k|^2}{\delta^2} + 2|\eta_\theta| \frac{|a_{l_k m_k}^k|}{\delta} \leq \frac{|a_{l_k m_k}^k|^2}{\delta^2} + \frac{2|\eta_\theta| \sqrt{W(0, k)}}{\delta} \end{aligned} \quad (85)$$

Therefore

$$\sum_{k=1}^{n_t(n_t-1)/2} \sin^2(\hat{\theta}_k^J) \leq \sum_{k=1}^{n_t(n_t-1)/2} \left(\frac{|a_{l_k m_k}^k|^2}{\delta^2} + \frac{2|\eta_\theta| \sqrt{W(0, k)}}{\delta} \right) = \frac{1}{\delta^2} W(0, k) + \frac{\eta_\theta (n_t^2 - n_t)}{\delta} \sqrt{W(0, k)} \quad (86)$$

By substituting (86) into (80) one obtains

$$P_{c_{n_t-1}}^2 \leq W(0, 0) \left(\frac{1}{\delta^2} W(0, 0) + \frac{(n_t^2 - n_t)|\eta_\theta|}{\delta} \sqrt{W(0, k)} \right) + \frac{|\epsilon|^2}{2} (n_t^2 - n_t), \quad (87)$$

It remains to relate η_θ in (84) to the accuracy of the line search η . Recall that the calculation of $\hat{\theta}_k^J$ relies on the calculation of $\hat{\phi}_k^J$. Thus, as a result of the finite accuracy of the line searches, η_θ depends on η_ϕ as

well, as we now show. Form the proof of [4, Theorem 2] we know that if an accurate line search were invoked, it would produce $\hat{\phi}_k^J = -\angle a_{l,m}^k$. However, the actual line search yields $\hat{\phi}_k^J = -\angle a_{l,m}^k + \eta_\phi$, where $|\eta_\phi| \leq \eta$. Thus, θ_k is obtained by searching the minimum of a perturbed version of $S(\mathbf{A}_k, \mathbf{r}_{l,m}(\theta, \phi_k^J))$, i.e.

$$\tilde{S}(\mathbf{A}_k, \mathbf{r}_{l,m}(\theta, \hat{\phi}_k^J)) = h_k(\cos^2(\theta)a_{l,l}^k - \cos(\eta_\phi)\sin(2\theta)a_{l,m}^k + \sin^2(\theta)a_{m,m}^k) \quad (88)$$

We first assume that $a_{ll}^k \neq a_{mm}^k$. From the proof of [4, Theorem 2], the optimal value of θ is $\theta_k^J = \frac{1}{2} \tan^{-1}(p_k)$, where $x_k = \frac{2|a_{l,m}^k|}{a_{m,m}^k - a_{l,l}^k}$. If one takes into consideration the non-optimality of the line-search which obtains $\hat{\phi}_k^J$ and ignores the non-optimality of the line search that obtains $\hat{\theta}_k^J$, then the minimizer of (88) would be $\theta_k^s = \frac{1}{2} \tan^{-1}(x_k \cos(\eta_\phi))$ and the difference $|\theta_k^J - \theta_k^s|$ is

$$|\theta_k^J - \theta_k^s| = \left| \frac{1}{2} \tan^{-1}(x_k \cos(\eta_\phi)) - \frac{1}{2} \tan^{-1}(x_k) \right| \leq \frac{|\eta_\phi \sin(\eta_\phi^*) p_k|}{\cos^2(\eta_\phi^*) p_k^2 + 1} \leq \eta_\phi^2 \frac{|x_k|}{\cos^2(\eta_\phi) x_k^2 + 1} \quad (89)$$

where $|\eta_\phi^*| \leq \eta_\phi$. It can be easily shown that $\frac{|x_k|}{\cos^2(\eta_\phi) x_k^2 + 1} \leq \frac{1}{|\cos(\eta_\phi)|}$, and because $\hat{\theta}_k^J = \theta_k^J + \theta_k^s - \theta_k^J + \eta_\phi$ and $|\eta_\phi| < \eta$, the accumulated effect of the finite accuracy of both line searches is bounded by $\eta_\theta \leq \eta + \frac{\eta^2}{|\cos(\eta)|}$. Assuming that η is sufficiently small, (e.g. $\eta \leq \pi/20$) we obtain

$$\eta_\theta \leq 6\eta/5 \quad (90)$$

By substituting (90) and (77) into (87) it follows that

$$P_{c_{n_t-1}}^2 \leq W(0,0) \left(\frac{1}{\delta^2} W(0,0) + \eta \frac{6(n_t^2 - n_r)}{5\delta} \sqrt{W(0,0)} \right) + (10 + 2\sqrt{2})(n_t^2 - n_r)\eta^2 \|\mathbf{G}\|^2 \quad (91)$$

Thus, as long as η is negligible with respect to $W(0,0)$, the BNSLA will have a quadratic convergence rate for \mathbf{G} that does not have multiple eigenvalues; i.e all eigenvalues are distinct. This is not sufficient since we are interested in a matrix \mathbf{G} $n_t - n_r$ with zero eigenvalues.

To extend the proof to the case where the matrix \mathbf{G} has $n_t - n_r$ zero eigenvalues and n_r distinct eigenvalues we use the following theorem:

Theorem 11 ([16] Theorem 9.5.1): Let \mathbf{A} be an $n_t \times n_t$ Hermitian matrix with eigenvalues $\{\lambda_l\}_{l=1}^{n_t}$ that satisfy

$$\lambda_1 \neq \lambda_2 \neq \dots \neq \lambda_{n_r} \neq \lambda_{n_r+1} = \lambda_{n_r+2} = \dots = \lambda_{n_t} = \lambda \quad (92)$$

Consider the following partition:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{B} \\ \mathbf{B} & \mathbf{A}_2 \end{bmatrix} \quad (93)$$

where \mathbf{A}_1 is $n_r \times n_r$ and \mathbf{A}_2 is $(n_t - n_r) \times (n_t - n_r)$ and let $\delta' > 0$. If $\|(\mathbf{A}_1 - \lambda \mathbf{I})^{-1}\| < 1/\delta'$, then

$$\|\mathbf{A}_2 - \lambda \mathbf{I}\| \leq \|\mathbf{B}\|^2/\delta' \quad (94)$$

To apply Theorem 11 to the modified O/BNSLA, we need to show that \mathbf{A}_k satisfies its conditions. This however is only satisfied in the next Jacobi sweep; i.e. Theorem 11 can be applied to \mathbf{A}_k with $k \geq m + 1$. To show this, note that (82) and (83) are satisfied by $\mathbf{A}_k, k \leq m$ for some permutations of the eigenvalues. Thus, due to the permutation in (22), $\mathbf{A}_k, k > m$ satisfies (82) and (83), for the ordering of (92). For the rest of the proof, it is assumed that $k > m$. Let $\mathbf{A}_1^k, \mathbf{A}_2^k, \mathbf{B}^k$ be \mathbf{A}_k 's submatrices that correspond to the partition in (93). Recall that in our case, $\lambda = 0$, thus, (82) implies that

$$\|\mathbf{A}_1^k\| > \delta, \quad (95)$$

and also implies that $a_{ll}^k \geq 5\delta/2, \forall 0 \leq l \leq n_r$. Furthermore, by [19, Corollary 6.3.4]

$$|\lambda_l(\mathbf{A}_1^k) - a_{ll}^k| \leq \|\mathbf{A}_k\|_{\text{off}} \leq \delta/2 \quad (96)$$

Thus

$$\lambda_l(\mathbf{A}_1^k) > 0, \forall 0 \leq l \leq n_r \quad (97)$$

and therefore, the matrix \mathbf{A}_1^k is invertible, and from (95), it follows that $\|(\mathbf{A}_1^k)^{-1}\| \leq 1/\delta$, which enables us to apply Theorem 11 to obtain

$$\|\mathbf{A}_2^k\| \leq \|\mathbf{B}_k\|^2/\delta \quad (98)$$

To show that (98) leads to quadratic convergence, one must show that the affiliation of the diagonal entries in the upper $n_r \times n_r$ -block of \mathbf{A}_k remains unchanged and that the eigenvalue that correspond are arranged in decreasing order, i.e.

$$l = \arg \min_{1 \leq m \leq n_t} |\lambda_l - a_{mm}^k| = \arg \min_{1 \leq m \leq n_t} |\lambda_l - a_{mm}^{k+1}|, \forall l \in \{1, \dots, n_r\} \quad (99)$$

and

$$\lambda_l \geq \lambda_m, \forall l \leq m \quad (100)$$

To show (99), note that

$$\left| a_{l_k, l_k}^k - a_{m_k, m_k}^k \right|^2 \leq \sin^2(\theta_k) \left(2 \cos(\theta_k) a_{l_k, m_k}^k \cos(\phi_k - \angle a_{l_k, m_k}^k) + \sin(\theta_k) (a_{l_k, l_k}^k - a_{m_k, m_k}^k) \right)^2 \quad (101)$$

and that for every θ_k such that $l_k \leq n_r$, (85) is satisfied. Thus

$$\begin{aligned} \left| a_{l_k, l_k}^k - a_{l_k, l_k}^{k+1} \right|^2 &\leq \sin^2(\theta_k) \left(a_{l_k, m_k}^k + \sin(\theta_k) (a_{l_k, l_k}^k - a_{m_k, m_k}^k) \right)^2 \\ &\leq \sin^2(\theta_k) \left(a_{l_k, m_k}^k + \left((a_{l_k, m_k}^k)^2 / \delta^2 + 2\eta_\theta a_{l_k, m_k}^k / \delta \right)^{1/2} \delta \right)^2 \\ &\leq \sin^2(\theta_k) \left(a_{l_k, m_k}^k + \left((a_{l_k, m_k}^k)^2 + 2\delta\eta_\theta a_{l_k, m_k}^k \right)^{1/2} \right)^2 \leq \sin^2(\theta_k) \left(a_{l_k, m_k}^k + (\delta^2/4 + 2\delta\eta_\theta\delta/2)^{1/2} \right)^2 \\ &\leq \sin^2(\theta_k) \left(a_{l_k, m_k}^k + \delta(1/4 + \eta_\theta)^{1/2} \right)^2 \leq \frac{\delta^2}{4} (1 + 4\eta_\theta) \left(1/2 + \sqrt{1/2 + \eta_\theta} \right)^2 \end{aligned} \quad (102)$$

By restricting $\eta \leq 1/100$ and considering (90) it follows that

$$|a_{l_k, l_k}^k - a_{l_k, l_k}^{k+1}| \leq 0.65\delta \quad (103)$$

which establishes (99).

Now that (99) is established, (100) immediately follows and for every l, m such that $l \neq m$ and $1 \leq l \leq n_r$, we have $|a_{ll}^k - a_{mm}^k| \geq \delta$. And (80) can be written as

$$\begin{aligned} P_{c_{n_r}+m} &\leq \sum_{l=1}^{n_t-1} Z(l, c_{n_r} + m) + |\epsilon^2| \sum_{l=1}^{n_r} (n_t - l) \\ &\leq W(0, m) \sum_{j=1}^{c_{n_r}} \sin^2(\theta_j) + \sum_{l=n_r+1}^{n_t-1} Z(l, c_{n_r} + m) + |\epsilon^2| \sum_{l=1}^{n_r} (n_t - l) \end{aligned} \quad (104)$$

Recall that $|\epsilon|^2 \leq \max_k u(\theta_k, \phi_k)$ where $u_1(\theta_k, \phi_k)$ and $u_2(\theta_k, \phi_k)$ are defined in (70) and (71). From (70) and (75) $u(\theta_k, \phi_k) \leq 4\eta^2(5P_k^2 + 4P_k\|\mathbf{G}\| + \|\mathbf{G}\|^2)$. Because $|a_{l_k l_k}^k - a_{m_k m_k}^k| \geq \delta$ for $m < k \leq m + c_{n_r}$, (86) is satisfied and similar to (91) we obtain

$$\begin{aligned} P_{c_{n_r}+m} &\leq W(0, m) \left(\frac{1}{\delta^2} W(0, m) + \eta \frac{(n_t^2 - n_r)}{\delta} \sqrt{W(0, k)} \right) \\ &\quad + 2(2n_t n_r - n_r^2 - n_r) \eta^2 (5W(0, m) + 4\sqrt{W(0, m)}\|\mathbf{G}\| + \|\mathbf{G}\|^2) + \sum_{l=n_r+1}^{n_t-1} Z(l, c_{n_r} + m) \end{aligned} \quad (105)$$

It remains to bound the term $\sum_{l=c_{n_r}+1}^{n_t-1} Z(l, c_{n_r} + m)$. Note that for every θ_k such that $m < k \leq c_{n_r} + m$, (85) is satisfied. Let $Q = \{(l, m) : 1 \leq l \leq n_r < m \leq n_t\}$. Note that for every k such that $(l_k, m_k) \in Q$, $a_{l_k l_k}^k$ and $a_{m_k m_k}^k$ are located in \mathbf{A}_1^k and \mathbf{A}_2^k respectively. Thus,

$$\begin{aligned} |a_{q, m_k}^{k+1}|^2 &\leq |a_{q, m_k}^k|^2 + \sin^2(\theta_k) |a_{l_k, q}^k|^2, \text{ for } n_r < q < m_k \\ |a_{m_k, q}^{k+1}|^2 &\leq |a_{m_k, q}^k|^2 + \sin^2(\theta_k) |a_{l_k, q}^k|^2, \text{ for } m_k < q \leq n_t \end{aligned} \quad (106)$$

and from (85)

$$\begin{aligned} |a_{m_k, q}^{k+1}|^2 &\leq |a_{m_k, q}^k|^2 + \left(\frac{|a_{l_k, m_k}^k|^2}{\delta^2} + 2\eta_\theta \frac{|a_{l_k, m_k}^k|}{\delta} \right) |a_{l_k, q}^k|^2 \text{ for } m_k < q \leq n_t \\ |a_{q, m_k}^{k+1}|^2 &\leq |a_{q, m_k}^k|^2 + \left(\frac{|a_{l_k, m_k}^k|^2}{\delta^2} + 2\eta_\theta \frac{|a_{l_k, m_k}^k|}{\delta} \right) |a_{l_k, q}^k|^2, \text{ for } n_r < q < m_k \end{aligned} \quad (107)$$

These can be bounded by

$$|a_{m_k, q}^{k+1}|^2, |a_{q, m_k}^{k+1}|^2 \leq W^2(0, m) \left(1 + \frac{1}{\delta^2}\right) + \frac{2\eta_\theta}{\delta} W^{3/2}(0, m) \quad (108)$$

Thus, for every $k \in \{m, \dots, m + c_{n_r}\}$,

$$\sum_{l=n_r+1}^{n_t-1} Z(l, c_{n_r} + m) = \sum_{q=n_r+1}^{n_t-1} \sum_{t=q+1}^{n_t} |a_{q, t}^k|^2 \leq O\left(\left(\frac{W(0, 0+m)}{\delta}\right)^2\right) + O\left(\left(\frac{\eta_\theta W^{3/2}(0, 0+m)}{\delta}\right)\right) \quad (109)$$

This, together with (105) and (90) show

$$P_{c_{n_r}}^2 \leq O\left(\left(\frac{W(0, 0+m)}{\delta}\right)^2\right) + O\left(\left(\frac{\eta W^{3/2}(0, 0+m)}{\delta}\right)\right) + O\left(\left(\frac{\eta^2 W^{1/2}(0, 0+m)}{\delta}\right)\right) + 2(n_t^2 - n_t) \eta^2 \|\mathbf{G}\|^2 \quad (110)$$

Since P_k is a decreasing sequence, the desired result follows. \square

APPENDIX C

PROOF OF THEOREM 5

We first prove the theorem for the case where the non-clustered eigenvalues are the largest; i.e., $\lambda_i \geq \lambda_{i+1} + \delta_c$ and $\lambda_i - \lambda \geq \delta_c$ for $i = 1, \dots, n_r - v$. Note that $\lambda_i = \lambda + \xi_{i-n_r-v}$ for $i \in L_2 = \{n_r - v + 1, \dots, n_r\}$ and $\lambda_i = 0$ for $i = n_r + 1, \dots, n_t$. Without loss of generality, we assume that $W(0, 0) \leq \delta_c^2/8$ where $W(k, l)$ is defined in (37). Let $\mathbf{V}_k \Lambda \mathbf{V}_k^* = \mathbf{A}_k$ be \mathbf{A}_k 's EVD, and let $\tilde{\mathbf{A}}^k = \mathbf{V}_k \tilde{\Lambda} \mathbf{V}_k^*$, $\hat{\mathbf{A}}^k = \mathbf{V}_k \hat{\Lambda} \mathbf{V}_k^*$ where

$$\begin{aligned} \tilde{\Lambda} &= \text{diag}(\lambda_1, \dots, \lambda_{n_r-v}, \underbrace{\lambda \cdots \lambda}_v, \underbrace{0 \cdots 0}_{n_t-v-n_r}) \\ \hat{\Lambda} &= \text{diag}(\underbrace{0 \cdots 0}_{n_r-v}, \xi_1, \dots, \xi_v, \underbrace{0 \cdots 0}_{n_t-n_r-v},) \end{aligned} \quad (111)$$

Let $L_1 = \{1, \dots, n_r - v\}$, $L_3 = L \setminus (L_1 \cup L_2)$ and $L_s = (L_1 \times L) \cup (L_2 \times L_3)$, $L_c = L_s \cap \{(l, m) : l < m\}$. By combining (81) and the condition $P_k^2 < \delta_c^2/8$, it follows that (82) and (83) hold for $\delta = \delta_c$. Thus, due to the permutation in (22), the inequalities (82) and (83) are satisfied for $\mathbf{A}_k, k > m, \forall (l, m) \in L_c$ and $\delta = \delta_c$. In the rest of the proof, we assume that $k > m$. Because $|a_{ll}^k - a_{mm}^k| < \delta_c, \forall (l, m) \in L_c$,

\mathbf{A}_k can be partitioned as

$$\mathbf{A}_k = \begin{bmatrix} \mathbf{A}_{11}^k & \mathbf{A}_{12}^k & \mathbf{A}_{13}^k \\ \mathbf{A}_{21}^k & \mathbf{A}_{22}^k & \mathbf{A}_{23}^k \\ \mathbf{A}_{31}^k & \mathbf{A}_{32}^k & \mathbf{A}_{33}^k \end{bmatrix} \quad (112)$$

where $\mathbf{A}_{22}^k \in \mathbb{C}^{(n_r-v) \times (n_r-v)}$ and $\mathbf{A}_{33}^k \in \mathbb{C}^{v \times v}$. The idea behind this partition is that the diagonal entries of \mathbf{A}_{11}^k are separated by more than δ_c , and two diagonal entries such that each belongs to a different diagonal block (i.e. $\mathbf{A}_{11}, \mathbf{A}_{22}, \mathbf{A}_{33}$) are also separated by more than δ_c . Now it is possible to use [14, Lemma 2.3] which asserts that

$$\|\mathbf{A}_{ll}^k\|_{\text{off}} \leq \frac{P_k^2}{2\delta_c}, \text{ for } l = 2, 3. \quad (113)$$

where $\|\mathbf{A}_{ll}^k\|_{\text{off}}$ is the sum of squares of \mathbf{A}_{ll}^k 's off diagonal entries.

To show that (113) establishes (29) we first show that the affiliation of the diagonal entries in the upper \mathbf{A}_{11}^k -block remains unchanged and that no diagonal entry leaves the \mathbf{A}_{22}^k and \mathbf{A}_{33}^k blocks. To be precise, for $i = 1, 2, 3$

$$R_{k+1}(v) \in L_i \text{ if } v \in L_i, \text{ and } R_k(v) = v \text{ if } v \in L_1, \quad (114)$$

where

$$R_k(v) = \arg \min_{l \in L} |\lambda_v - a_{ll}^k| \quad (115)$$

This follows from (101) and because for every k such that $(l_k, m_k) \in L_c$, (85) is satisfied with replacing δ by δ_c . Thus, similar to (102), for every k such that $(l_k, m_k) \in L_s$ $\left| a_{l_k, l_k}^k - a_{l_k, l_k}^{k+1} \right|^2 \leq \frac{\delta_c^2}{4} (1 + 4\eta_\theta) \left(1/2 + \sqrt{1/2 + \eta_\theta} \right)^2$. By taking $\eta \leq 1/100$ and considering (90) it follows that $|a_{l_k, l_k}^k - a_{l_k, l_k}^{k+1}| \leq 0.65\delta_c$, which establishes (115); and therefore, for every $(l, m) \in L_s$, $|a_{ll}^k - a_{mm}^k| \geq \delta_c$.

Similar to the derivation of (105),

$$\begin{aligned} P_{c_{n_t-v-r}+m} &\leq W(0, m) \left(\frac{1}{\delta_c^2} W(0, m) + \eta \frac{n_t^2 - n_t}{\delta_c} \sqrt{W(0, k)} \right) \\ &\quad + 4c_{n_t-v-r} \eta^2 (5W(0, m) + 4\sqrt{W(0, m)} \|\mathbf{G}\| + \|\mathbf{G}\|^2) + \sum_{l=n_t+v+r+1}^{n_t-1} Z(l, c_{n_t-v-r} + m) \end{aligned} \quad (116)$$

and similar to the derivation of (110), we obtain

$$P_{c_{n_t-v-r}+m}^2 \leq O \left(\left(\frac{W(0, m)}{\delta_c} \right)^2 \right) + O \left(\left(\frac{\eta W^{3/2}(0, m)}{\delta_c} \right) \right) + O \left(\left(\frac{\eta^2 W^{1/2}(0, m)}{\delta_c} \right) \right) + 2(n_t^2 - n_t) \eta^2 \|\mathbf{G}\|^2$$

Since P_k is a decreasing sequence, the desired result follows. \square

APPENDIX D

PROOF OF COROLLARY 7

The corollary follows from Theorem 3 and from the following proposition:

Proposition 12: Let $b > 0, 0 < \rho < 1$ and let a_n be a non-negative sequence that satisfies

$$a_{n+1} \leq \rho a_n + b, \quad \forall n \in \mathbb{N}, \quad (117)$$

then,

$$\limsup_n a_n \leq \frac{b}{1-\rho} \quad (118)$$

Proof: We first assume that for some $n \in \mathbb{N}$ $a_n \geq \frac{b}{1-\rho}$. In this case we have $a_{n+1} \leq a_n$ which means that a_n is a monotonic decreasing sequence as long as $a_n \geq \frac{b}{1-\rho}$. In the case where $a_n < \frac{b}{1-\rho}$ we have $a_{n+1} < \frac{b}{1-\rho}$. These mean that either a_n converges to a limit $\xi > \frac{b}{1-\rho}$, or that it satisfies (118). Assume that the previous statement is true, then for every $\epsilon > 0$, there exists $n_\epsilon \in \mathbb{N}$ such that $\xi - \epsilon \leq a_n \leq \xi + \epsilon, \forall n > n_\epsilon$. By substituting it into (117), i.e., substituting $\xi - \epsilon$ for a_{n+1} and $\xi + \epsilon$ for a_n it follows that for every $\epsilon > 0$, $\xi(1-\rho) \leq b + \epsilon(1+\rho)$. This is equivalent to $\xi \leq \frac{b}{1-\rho} + \frac{\epsilon(1+\rho)}{(1-\rho)}, \forall \epsilon > 0$ which is a contradiction. \square

REFERENCES

- [1] R. Zhang and Y.-C. Liang, "Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, pp. 88–102, 2008.
- [2] G. Scutari, D. Palomar, and S. Barbarossa, "Cognitive MIMO radio," *IEEE Signal Process. Mag.*, vol. 25, pp. 46–59, November 2008.
- [3] G. Scutari and D. Palomar, "MIMO cognitive radio: A game theoretical approach," *IEEE Trans. Signal Process.*, vol. 58, pp. 761–780, Feb. 2010.
- [4] Y. Noam and A. Goldsmith, "Blind null-space learning for MIMO underlay cognitive radio with primary user interference adaptation," *to appear, IEEE Trans. Wireless Commun.*
- [5] H. Yi, "Nullspace-based secondary joint transceiver scheme for cognitive radio MIMO networks using second-order statistics," in *IEEE Int. Conf. Commun.*, pp. 1–5, May 2010.
- [6] Z. Chen, C.-X. Wang, X. Hong, J. S. Thompson, S. A. Vorobyov, F. Zhao, H. Xiao, and X. Ge, "Interference mitigation for cognitive radio MIMO systems based on practical precoding," *Arxiv preprint arXiv:1104.4155*, vol. abs/1104.4155, 2011.
- [7] F. Gao, R. Zhang, Y. Liang, and X. Wang, "Design of learning-based MIMO cognitive radio systems," *IEEE Trans. Vehic. Tech.*, vol. 59, pp. 1707–1720, May 2010.
- [8] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins University Press, 3 ed., 1996.

- [9] G. E. Forsythe and P. Henrici, "The cyclic Jacobi method for computing the principal values of a complex matrix," *Trans. American Math. Soc.*, vol. 94, no. 1, pp. 1–23, 1960.
- [10] P. Henrici and K. Zimmermann, "An estimate for the norms of certain cyclic Jacobi operators," *Linear Algebra and Its Applications*, vol. 1, no. 4, pp. 489–501, 1968.
- [11] K. V. Fernando *Numerische Mathematik*, vol. 56, pp. 73–91, 1989.
- [12] P. Henrici, "On the speed of convergence of Cyclic and Quasicyclic Jacobi methods for computing eigenvalues of Hermitian matrices," *J. Soc. Ind. App. Math.*, pp. 144–162, 1958.
- [13] J. H. Wilkinson, "Note on the quadratic convergence of the Cyclic Jacobi process," *Numerische Mathematik*, vol. 4, pp. 296–300, 1962. 10.1007/BF01386321.
- [14] V. Hari, "On sharp quadratic convergence bounds for the serial Jacobi methods," *Numerische Mathematik*, vol. 60, pp. 375–406, 1991.
- [15] R. P. Brent and F. T. Luk, "The solution of Singular-Value and Symmetric Eigenvalue Problems on multiprocessor arrays," vol. 6, no. 1, pp. 69–84, 1985.
- [16] B. N. Parlett, *The Symmetric Eigenvalue Problem*. Soc. Industrial Math., 1998.
- [17] G. H. Golub and H. A. van der Vorst, "Eigenvalue computation in the 20th century," *Journal of Computational and Applied Mathematics*, vol. 123, no. 1-2, pp. 35 – 65, 2000.
- [18] Y. Noam and A. J. Goldsmith, "Exploiting spatial degrees of freedom in MIMO cognitive radio systems," *IEEE International Conference on Communications (ICC)*, 2012.
- [19] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1987.
- [20] C. Xiao, Y. R. Zheng, and N. Beaulieu, "Novel sum-of-sinusoids simulation models for rayleigh and Rician fading channels," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3667 –3679, 2006.
- [21] A. Liou, K. Sivertsen, and D. Michelson, "Characterization of time variation on 1.9 GHz fixed wireless channels in suburban macrocell environments," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 3975–3979, 2009.